Statistical Aspects of Bioequivalence Assessment in the Pharmaceutical Industry

Scott Daniel Patterson, B.A., M.Sc., M.Phil.
Statistical Sciences, University College London
United Kingdom
Dissertation submission for the Degree of Doctor of Philosophy

Supervised by
Byron Jones
and
Stephen Senn

July 4, 2003

Scott Daniel Patterson, B.A., M.Sc., M.Phil.
Title: Statistical Aspects of Bioequivalence Assessment in the Pharmaceutical Industry

Abstract

Since the early 1990's, average bioequivalence studies have served as the international standard for demonstrating that two formulations of drug product will provide the same therapeutic benefit and safety profile when used in the marketplace.

Population (PBE) and Individual (IBE) bioequivalence have been the subject of intense international debate since methods for their assessment were proposed in the late 1980's. Guidance has been proposed by the Food and Drug Administration of the United States government for the implementation of these techniques in the pioneer and generic pharmaceutical industries. As of the present time, no consensus among regulators, academia, and industry has been established. The need for more stringent population and individual bioequivalence has not been demonstrated, and it is known that the criteria proposed by FDA are actually *less* stringent under certain conditions.

The properties of method-of-moments and restricted maximum likelihood modelling in replicate designs will be explored in Chapter 2, and the application of these techniques in the assessment of average bioequivalence will be considered. Individual and population bioequivalence criteria in replicate cross-over designs will be explored in Chapters 3 and 4, respectively, and retrospective data analysis will be used to characterise the properties and behaviour of the metrics.

Simulation experiments will be conducted in Chapter 5 to address questions arising from the retrospective data analyses in Chapters 2 through 4. Additionally, simulation will be used to explore of a potential phenomenon known as 'bio-creep' - that is the transitivity of individual bioequivalence in practice.

Another bioequivalence problem is then considered to conclude the thesis; that of comparing rate and extent of exposure between differing ethnic groups as described in ICH-E5 (1998). The properties of the population bioequivalence metric and an alternative metric will be characterised in small and large samples from parallel group studies. Inference will be illustrated using data from a recent submission and simulation studies.

Contents

List of Figures

List of Tables

Acknowledgements: The author gratefully acknowledges the contributions and advice of Névine Zariffa, Vladimir Dragalin, Val Fedorov, Walter Hauck, Terry Hyslop, Darryl Downing, and the PhRMA Expert Panel on Bioequivalence in motivating some aspects of the work which has been reported in this thesis.

1 Introduction

The materials of this chapter were presented at the University College, London Journal Club (Patterson, 2000c), at the Barlett International Conference on Bioavailability and Bioequivalence (Patterson, 2001a), at the American Statistical Association, Philadelphia Chapter Meeting (Patterson, 2001b), and at the Drug Information Association meeting (Jones and Patterson, 2002). The materials were published at the request of the editor of the *Indian Journal of Pharmaceutical Sciences* (Patterson, 2001c-d).

1.1 Definition of Bioequivalence studies

Bioequivalence studies are performed to demonstrate that drug products are similar to each other in terms of their therapeutic benefit (efficacy) and non-therapeutic side effects (safety) and as such play a key and pivotal role in the drug development process. They are primarily utilized in the study of solid oral dosage forms (i.e. drugs administered as a tablet or capsule when ingested), and this thesis will confine itself to discussion of these type of drug products.

The study of what happens to such a dose of drug substance when it is released into the human body is known as pharmacokinetics (PK). Following oral administration, the drug is held to undergo four 'stages' prior to being completely eliminated from the body, known as **ADME**: **A**bsorption (uptake by the body through the mouth, throat, stomach, and small/large intestine), **D**istribution (how the drug substance is carried by the body through the blood to its site of action), **M**etabolism (how the body breaks the drug substance into by-products), and **E**limination (how the body disperses the drug product). The study of PK is typically referred to as 'what the body does to the drug'.

Once a drug is ingested, the substance (or active metabolite) passes through the blood and reaches its site of action; thereby provoking what is termed a pharmacodynamic (PD) response in the body. This PD response is regarded as a surrogate marker, which is indicative of subsequent therapeutic benefit, *e.g.* consider an insulin sensitiser in the treatment of diabetes. By causing the body to use its own insulin more effectively, the body is able to better regulate glucose levels (the surrogate marker), thereby improving the symptoms associated with the disease. It should be noted however that, at the same time, the drug (or again a metabolic by-product) may attach itself to a different site of action thereby provoking unwanted response or side effects, *e.g.*

weight gain in the treatment of diabetes. The study of pharmacodynamics is usually referred to as 'what the drug does to the body'. In combination, the study of dose, PK, and PD are held to determine the behaviour of a drug product (see Figure 1). These properties are characterized across a drug's clinical development.

Figure 1: Dose - Blood Concentration - Response Relationships; Sheiner et al. (1989)

Clinical development of a drug product, with the exception of only the most toxic products targeted for the treatment of cancer, initiates with the study of the drug product in normal healthy male volunteers (Phase I). These studies are typically small, well-controlled, data-intensive, dose escalating, and placebo-controlled. In this stage of development, the primary objective of a clinical study is to determine a safe range of doses and dosing regimens (e.g. once-a-day or twice-a-day) for later dosing in studies involving patients with the disease state under study. Dose and dosing regimen are examined with respect to their impact on the PK of the drug product, and additionally, should biomarker or surrogate markers be present to characterize the pharmacodynamic activity of the drug in normal healthy volunteers, these data are characterized relative to dose and/or PK levels. It is unlikely to be the case at this stage of development (again with the notable exception of those drug products being studied for oncology indications) that PD or therapeutic response can be examined.

Clinical studies in Phase II establish the minimum starting and maximum effective dose as well as the maximum tolerated dose in patients with the disease state using pharmacodynamic surrogate or markers of therapeutic response. Dose titration and the length of time needed to see an effect (desirable or undesirable) are also established. In these studies, models relating dose to PK and to PD are developed to understand the mechanism of the drug's action and to search for relevant covariates (e.g. age or gender) to control later Phase II or Phase III confirmatory trial designs (International Conference on Harmonization Guidance ICHE4, 1994). Study designs and analysis procedures for dose finding trials are described in more detail in Patterson et al. (2000d) and fall beyond the scope of this thesis.

Once a dose or set of efficacious doses are chosen from Phase II trials, confirmatory trials are subsequently performed to support regulatory acceptance of the risk relative to benefit in clinical use of the compound in large numbers of patients with the disease under study.

In parallel with Phase II studies in patients, *clinical pharmacology* studies are conducted in normal healthy volunteers to refine knowledge of the PK and PD of the compound. A list of potential studies is provided in Table 1.

Table 1: Studies during Phases II and III Typically Performed in a Clinical Pharmacology Package for Submission to Regulatory Authorities

Phase II	Phase III
Relative Bioavailability of Confirmatory Trials to Phase I Formulation	Bioequivalence of Confirmatory Trials to Commercial Formulation
Effect of Age and Gender on PK of Confirmatory Trials Formulation	Effect of Food on PK of Commercial Formulation
Effect of Food on PK of Confirmatory Trials Formulation	Effect of Renal Disease on PK of Commercial Formulation
IV Dose Finding Study	Dose-proportionality of Commercial Formulation
Drug Interactions on PK of Confirmatory Trials Formulation	ADME using IV and Oral radiolabelled drug
	Liver Disease on PK of Commercial Formulation
	Drug Interactions on PK of Commercial Formulation

One such study in Phase II, known as a relative bioavailability study, is performed to facilitate

formulation selection. These studies are primarily used by pharmaceutical sponsors of new drug entities to ensure that the formulation to be used in Phase II or in confirmatory trials is sufficiently similar to that used in Phase I drug development. Many clinical pharmacology studies are also performed in parallel with confirmatory trials in Phase III; however, the most important study for discussion in this thesis is the bioequivalence study to demonstrate that the formulation used in Phase III clinical trials is sufficiently similar to the final commercial formulation to be marketed following approval. These studies are primarily used by pharmaceutical sponsors of new drug entities who have conducted pivotal efficacy trials with a specific formulation of a drug therapy but need or want market access for a more commercially suitable formulation. These studies can be viewed as providing reassurance to regulators that the formulation to be marketed is the same as that used in the clinical confirmatory trials without the need to repeat the development programme or to perform a therapeutic equivalence study in patients with clinical endpoints (Huque et al., 1989).

Bioequivalence studies are conducted to meet pre-set regulatory standards and are the focus of this thesis. Relative bioavailability studies in contrast are conducted by companies developing new chemical entities for internal company reassurance that formulation changes during Phases I and II have not impacted the rate and extent of bioavailability of the compound and are not held to a strict acceptance standard. While methods of analysis (see Sections 1.3-1.5) may be applied in these studies, often small changes in PK are acceptable to sponsors provided negligible impact is expected in the PD response in clinic. Attention in this thesis will now be confined to bioequivalence studies (i.e. studies required for regulatory acceptance of a drug product as equivalent to that shown as efficacious and safe in the clinical development programme.)

One might ask, 'If the new and old formulations use exactly the same substance, i.e. are pharmaceutically equivalent (Benet, 1999), why do we do these studies at all?' Rate and extent of bioavailability in vivo can be drastically affected by very small changes in the constituent content of the formula, by small changes to the lining of the formula, and by compaction into tablet (versus administration as a capsule), for example. Further discussion on this topic may be found in Levy (1995) and Balthasar (1999).

Cross-over study designs (Jones and Kenward, 1989; Senn, 1993; Senn, 2002) are typically used to study bioequivalence. These studies will be discussed at length in Section 1.2 but

are discussed briefly here. Such designs are typically conducted in normal healthy volunteer subjects. Each subject is administered formulations (**T**est or **R**eference formulations in the example) in a sequence of treatments with each administration separated by a washout period appropriate to the drug under study, see Table 2. In these studies, subjects provide data on multiple separate sessions separated by adequate washout to avoid residual drug concentrations from the previous occasion. These studies are usually performed at the maximum tolerated, single oral dose as regulators have deemed a single oral dose to be most sensitive to changes in bioavailability (el-Tahtawy et al., 1998; FDA Guidance 1992-2002). In general, subjects receive each drug formulation at least once over the duration of the study, and plasma samples are collected following each administration to derive a concentration versus time profile, see Figure 2.

Table 2: A Two Period Cross-over Study Design with Test and Reference Formulations

Subject	Period 1	Wash-out Period of 5 half-lives	Period 2
001	T	.....	R
002	R	.....	T
003	R	.....	T
		.....	

Summary measures (Rowland and Tozer, 1980) for the plasma concentration versus time curve are derived as:

* AUC(0-t) (Area under the curve from time zero to t where t is the time of last quantifiable concentration),

* Cmax (maximal concentration)

* Tmax (time of maximal concentration),

* T1/2 (half-life of drug substance), and

$$AUC(0-inf) = AUC(0-t) + \frac{C_t}{\lambda} \tag{1}$$

where C_t is the concentration at time t and λ is -2.303 times the slope of the terminal phase of the $\log_e$-concentration time curve. See Steinijans et al. (1995) for other summary measures. More details of techniques used in derivation of AUC may be found in Yeh and Kwan (1978).

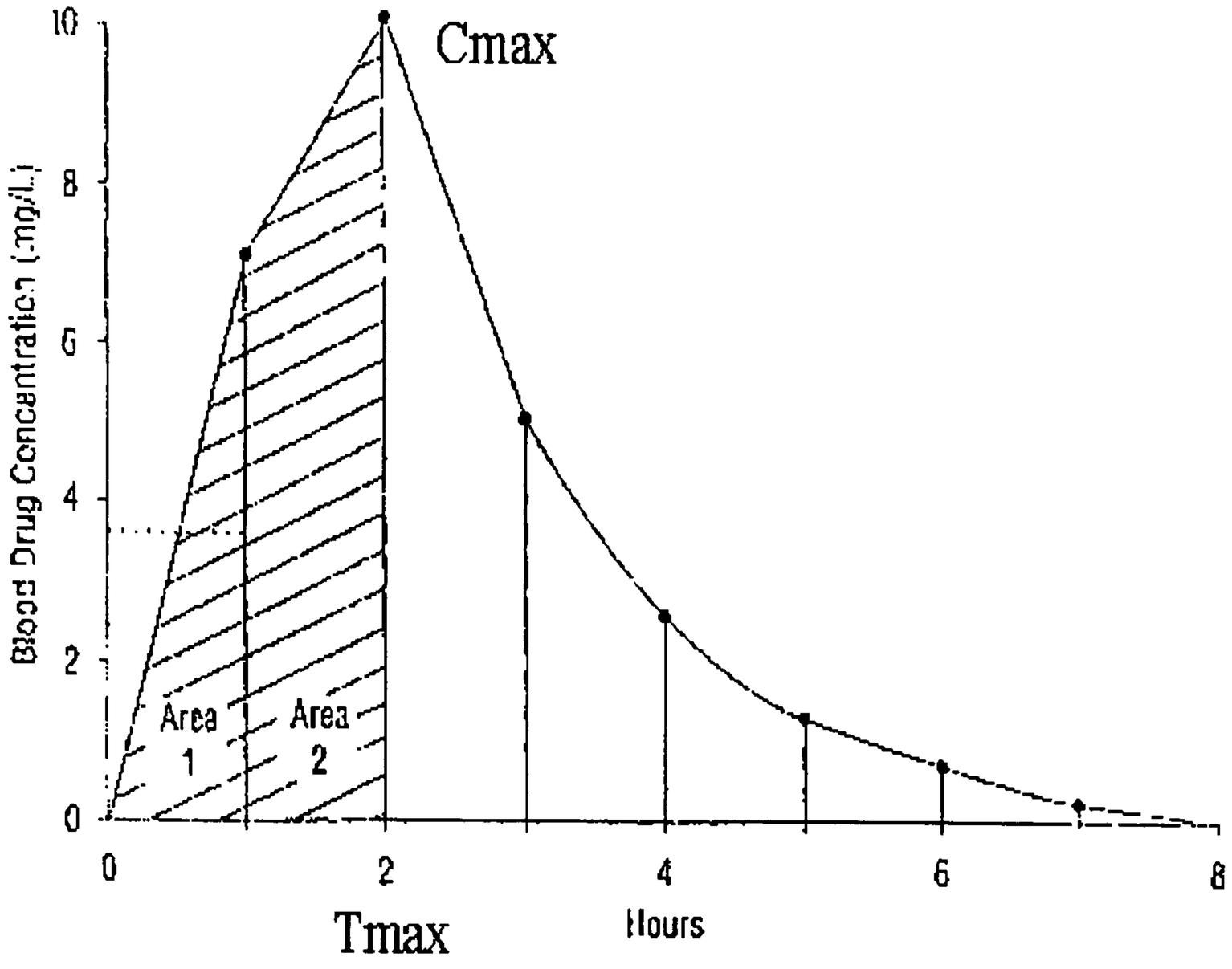


Figure 2: Example - Plasma Concentration (ng/mL) versus Time (h)

To demonstrate equivalence in plasma concentration profiles, *rate* and *extent* of bioavailability of the drug substance in plasma must be sufficiently similar so as to remove all doubt that exposure of the body to the drug substance is the same between formulations (Benet, 1999). For this purpose, Cmax (rate) and AUC (extent) are typically used as summary measures for the plasma concentration curves and are required to be demonstrated as equivalent under preset decision rules to compel regulatory approval. The other pharmacokinetic endpoints provide supporting information but do not directly impact approvability of the new formulation. Some papers would appear to indicate that AUC and Cmax are not sufficient to prove bioequivalence (Resigno and Powers, 1998; Steinijans et al., 1995; Lacey et al., 1995); however, international regulatory authorities have depended on these endpoints since the early 1990s. Pharmacodynamic data or safety data may be required for some drug products (for example, see Marzo et al., 2000).

Statistical moment theory denotes AUC as the first moment of the concentration time curve (Yamaoka, Nakagawa, Uno, 1978). This measure is held by international regulators (Cartwright, 1991; Herchuelz, 1996; el-Tahtawy et al., 1998) to be a standard measure for extent of bioavailability.

Cmax as a measure of rate of bioavailability has been found to be confounded with extent of bioavailability in studies (Basson et al., 1998) and is known to not characterize the rate of bioavailability particularly well (Cartwright, 1991). Given its dependence on the a priori choice of sampling scheme, it is known to be more variable than AUC and is sometimes problematic in the assessment of bioequivalence (Tsang et al., 1996; Buice et al., 1996). Regardless of this however, Cmax has been held to be more reliable in the eyes of regulators than several alternatives (Bois et al., 1994).

Other measures of rate of absorption have been proposed in the literature such as Direct Curve Metrics (Marston and Polli, 1997) and Cmax/AUC (Endrenyi et al., 1991), and 'indirect' metrics (Ring et al., 2000). However, simulation based assessment of alternatives has demonstrated such measures to be less desirable than the use of Cmax to date (Tozer et al., 1996, Tozer and Hauck, 1997). Recent work in alternative measures of absorption rate such as Partial AUCs (Endrenyi et al., 1998a) is ongoing in response to workshop and regulatory considerations (Patnaik et al., 1997; Shah et al., 1996) but have yet to be accepted as useful measures in bioequivalence assessment (Barrett et al., 2000). Cmax thus seems to be held as the 'least evil' measure available at present for rate of bioavailability (el-Tahtawy et al., 1998).

Endpoints are typically subjected to analysis separately. Multiplicity adjustment in analysis is not required for regulatory approval under preset 'joint decision rules' (Hauck et al., 1995).

Pharmacokinetic endpoints AUC and Cmax are generally held to be log-normally distributed (Westlake, 1979; Midha et al., 1993; Lacey et al., 1997; Julious and Debarnot, 2000). If X is log-normally distributed with mean $\exp(\mu + (1/2)\sigma^2)$ and variance $\exp(2\mu + \sigma^2)(\exp(\sigma^2)\text{-}1)$, then $Y=(\log_e(X))$ is known to be normally distributed with mean μ and variance σ^2 (Crow and Shimizu, Section 1.1, 1988; Crowder et al., Ch 2.5, 1991). Thus, typically, in the analysis of PK studies, AUC and Cmax are simply $\log_e$-transformed and subjected to analysis on the normal scale. Differences constructed on the normal scale are exponentiated and expressed as the ratio of two geometric means. Analysis will be discussed in detail in Sections 1.2 and 1.3 and is typically implemented based on the use of mixed models (Patterson and Thompson, 1971; Laird and Ware, 1982; Jones and Kenward, 1989; Milliken and Johnson, 1992; Senn, 1993; Vonesh and Chinchilli, 1997; Senn, 2002).

Bioequivalence studies must also be performed following substantial post-marketing formu-

lation alteration. They are also used by the generic pharmaceutical industry to gain market access for generic formulations of established drug therapies when the patent of the sponsor's formulation expires or when the original sponsors themselves perform a formulation change (for instance, change the site of manufacture) following approval.

Multiple companies may produce and market similar formulations to the original marketed product following patent expiration, provided they can demonstrate bioequivalence to the original product. *Generic substitution* has thus provided a means of providing the market with inexpensive, efficacious, and safe drug products without the need to repeat an entire clinical and clinical pharmacology development package following patent expiration.

The history of bioequivalence testing and generation of decision rules for regulatory approval are now examined in greater detail.

1.2 History of Bioequivalence in the 1970's to Early 1980's

Idiosyncratic reports of therapeutic failure and/or undesirable side effects are to be expected when a drug is administered to a large human population in the marketplace. Clinical trials used to register a drug for admittance to the marketplace can, in general, sample only a small subset of those people who will eventually use the product in their therapy, and not all patients in the market will derive benefit and have a completely safe experience while taking any given product. Patient risk of such an occurrence, however, should be minimised, or held at acceptable levels, as a matter of public policy.

In the late 1960s and 1970s, advances in chemical engineering increased the capabilities of the pharmaceutical industry to create inexpensive copies of patented drug products (since termed 'generics'). Following patent expiration, such new formulations could potentially be marketed with substantial profits for the producing company (Strom, 1987). Coupled with this possibility of increasing supply of the products in demand in the marketplace, this offered substantive benefit to public health (lower costs), and an explosive growth in the generic pharmaceutical industry was potentially to follow.

However, reports of failure of some generic drug products received a great deal of public attention in the United States following their documentation earlier in this time period (see Table 3), and it was concluded that development of a process and set of standards for market

access was necessary (Rheinstein, 1990) especially for those drugs with a narrow therapeutic index (drugs for which a small change in dose or rate/extent of bioavailability can result in large changes in response, see Ansbacher, 1990).

Table 3: Examples of Bioinequivalence (Calvert, 1996)

Image removed due to third party copyright

Further examples of bioinequivalence for carbamazepine (indicated for seizure control) may be found in Welty (1992).

Generic substitution was thus encouraged as part of the 1984 Drug Price Competition and Patent Term Restoration Act passed by the US Congress, since codified as 21 Code of Federal Regulations, Passage 320. FDA was authorised under this act to create an approval process for generic drug products, and subsequent to the act's passage, in 1985, FDA approved the marketing of one-hundred twenty-two generic equivalent formulations for products in the marketplace with expired patents (Strom, 1987).

The year following revealed increasing trends in market access for generic products (Strom, 1987). Approved drug products with therapeutic equivalence evaluations are listed by the US Food and Drug Administration in the *Orange Book* (Hare and Foster, 1990). For approval to market, FDA required a bioequivalence study for market access with prespecified decision rules for acceptability based on the data collected. Such studies were also required for extension of patent protection for innovators seeking to maintain market exclusivity (Hunt, 2000).

Study design and analysis procedures for such bioequivalence studies had been under consideration for some time by the time the act was passed among the statistical community (Metzler, 1974). Although statistical approaches to study design have historically concentrated on concise definition of the questions of interest to facilitate precise definition of sample size, randomisation, and control requirements (Hinkelmann and Kempthorne, 1994), the hypothesis of interest for the question of bioequivalence have proved to be elusive. However, in response to the practical

requirement for market access, a standard design has evolved.

In the 1960s through the 1980s, cross-over designs were a general topic of interest in clinical research (see Grizzle, 1965; Federer and Raghavarao, 1975; Hills and Armitage, 1979; Kershner and Federer, 1981; Freeman, 1989), and their statistical properties received much discussion and development. Cross-over designs are experiments where each experimental unit (the unit to which a treatment is administered) receive sequences of treatments (Jones and Kenward, 1989). Their use in the area of clinical research was deemed to be most appropriate (i.e. straightforward) when disease state was the same for any given patient (or subject) through the course of the study and when it could be assumed that carryover effects from treatment in the preceding period were negligible (Fleiss, 1986 and 1989). The use of such a design can save resources under certain conditions (Brown, 1980).

Comprehensive discussion of the design and analysis of cross-over trials may be found in Jones and Kenward (1989) and Senn (1993, 2002), and recent developments are reviewed in Jones and Deppe (2000). Discussion in this thesis will be confined to those cross-over studies typically used in the assessment of bioequivalence.

Randomized cross-over designs in normal healthy volunteers (of standard age and weight) have evolved as the design of choice for assessment of bioequivalence. Each normal healthy volunteer (hence termed subject) is administered each formulation in a predetermined, randomized sequence. Different formulations are thus administered to different subjects in each period, as in the design presented in Table 2. Administration of each formulation in each period is separated by a washout period sufficiently long to ensure non-quantifiable plasma concentrations at the beginning of the following period. Carryover effects in pharmacokinetic assessment are thus held to be negligible (confirmed by collection of a plasma sample prior to dosing in each period). Statistical testing procedures described in Jones and Kenward (1989) for the presence of first-order carryover (though sometimes the subject of intense debate; Senn, 1996, 1997) generally find carryover to be negligible (Zariffa et al., 2000; D'Angelo et al., 2001) allowing for straightforward implementation of the use the cross-over design and interpretation of the data in the assessment of bioequivalence. Such a study is said to be balanced if equal numbers of subjects receive each sequence.

Summary measurements such as AUC from a two-by-two cross-over trial may be modeled

using a random-intercept mixed modelling procedure accounting for each subject as their own control (Jones and Kenward, Ch 7, 1989; Milliken and Johnson, Ch 32, 1992). In bioequivalence studies, the following model for observations is commonly accepted for a randomized, two period, cross-over trial in normal healthy volunteers, under the assumption that carryover effects are negligible (or similar between formulations). Let y_{ijk} be the $\log_e$-transformed j-th period's observation ($j = 1, 2$) for the k-th subject ($k = 1, 2, \ldots, n_i$) in the i-th sequence group ($i = 1, 2$). Then

$$y_{ijk} = \lambda_i + (\mu + \nu_{k(i)}) + \pi_j + \tau_{d[i,j]} + \varepsilon_{ijk} \tag{2}$$

where μ is the grand mean,

λ_i, π_j, and $\tau_{d[i,j]}$ are fixed effects for sequence, period, and formulation,

$\nu_{k(i)}$ and ε_{ijk} are random effects which are independent with mean zero, $\mathrm{Var}(\nu_{k(i)})=\sigma_B^2$, the between-subject variance, and

$\mathrm{Var}(\varepsilon_{ijk})=\sigma_W^2$, the within-subject variance. This equation is usually expressed in practice as:

$$y_{ijk} = \lambda_i + \nu_{k(i)} + \pi_j + \mu_d + \varepsilon_{ijk} \tag{3}$$

where $d = (T, R)$. Analysis of data under $\log_e$-transformation is described in Box and Cox (1964).

For balanced designs ($n_1=n_2$) with no missing data, period effects are orthogonal to formulation effects. Note that homogeneity of test and reference product between-subject variance is assumed in such a random-intercept model, as is the homogeneity of within-subject variance for test and reference products. Observations between-subject are held to be independent (with covariance zero), and the variance-covariance structure for observations within a subject under these assumptions is compound symmetric. Under this model, the $Cov(y_{ijk}, y_{ij'k}) = E(\nu_{k(i)} + \varepsilon_{ijk})(\nu_{k(i)} + \varepsilon_{ij'k}) = E(\nu_{k(i)}^2) = Var(\nu_{k(i)}) = \sigma_B^2$ such that:

$$\rho = Corr(y_{ijk}, y_{ij'k}) = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} \tag{4}$$

for $j \neq j'$, and

$$\sigma^2 = Var(y_{ijk}) = \sigma_B^2 + \sigma_W^2 \tag{5}$$

Comparisons between the estimated means $\hat{\mu}_T$-$\hat{\mu}_R$ are thus normally-distributed with mean μ_T-μ_R and variance of $((\sigma_B^2 + \sigma_W^2) + (\sigma_B^2 + \sigma_W^2) - 2(\rho)(\sqrt{\sigma_B^2 + \sigma_W^2})(\sqrt{\sigma_B^2 + \sigma_W^2}))/n = 2(\sigma_W^2)/n$ in balanced two-period cross-over designs with no missing data and $n = n_1 + n_2$ subjects.

FDA initially proposed Decision Rules (sometimes referred to as uniformity requirements) to assess bioequivalence such as the 80/20 and 75/75 rule. Under these criteria, the study first (the 80/20 rule) must **not** have rejected the hypothesis that

$$H_0: \quad \mu_T = \mu_R \tag{6}$$

versus

$$H_1: \quad \mu_T \neq \mu_R \tag{7}$$

Additionally the study must have had sufficient number of subjects and low enough within-subject variance to have had eighty percent post-hoc power (probability of demonstrating bioequivalence under this decision rule when it is in fact bioequivalent) to detect a clinically important difference, usually defined to be $\ln 1.25$ on the log_e-scale (a twenty percent difference on the natural scale). For some additional products, see Haynes (1981), an additional requirement (the 75/75 rule) was defined such that seventy-five percent of subjects' individual ratios of test to reference must be greater than or equal to the value of 0.75 for bioequivalence to be demonstrated.

Criticisms of the 80/20 approach to bioequivalence are statistically obvious. Absence of evidence does not imply evidence of absence - i.e. statistical significance does not imply clinical significance, and vice versa (for more discussion see Altman and Bland, 1995 and Jones et al., 1996), and the statistical community had been aware, for some time, of better methods to test the hypothesis of equivalence of two treatments relative to a pre-set, clinically relevant goalpost.

Furthermore, the use of post-hoc power calculations is inappropriate in this context (Hoenig and Heisey, 2001).

On a practical level, Haynes (1981) established using simulation studies that the proposed 75/75 uniformity requirement was highly dependent on the magnitude of within-subject variation. Additionally, as can readily be seen from the above model (2), individual ratios are confounded with period effects. As these effects are known to frequently appear as significant in cross-over studies in normal healthy volunteers (Schuirmann, 1990), due for example to changes in assay procedures between periods, use of the 75/75 rule criteria for bioequivalence assessment was quickly observed to be inappropriate for a large variety of drug products and was dismissed from regulatory practice.

Cox (1967) related Fieller's theorem (1954) for the ratio of two normal distributed means to the conditional distributions used to obtain similar regions based on traditional Neyman-Pearson theory (for the testing of hypotheses; see also Locke, 1984). Alteration of the traditional hypothesis tested in clinical trials (equations 6 and 7, above), to a framework appropriate for equivalence testing, was introduced by Dunnett and Gent (1977). In this paper, Dunnett and Gent compared two binomial samples relative to a prespecified goalpost δ to assess equivalence of the responses to treatment. Westlake (1972-1979; for summary of work performed in the 1970's see Westlake, 1986) applied similar concepts to the analysis of bioequivalence trials. In brief, a bioequivalence study is conducted, and the confidence interval for the ratio of μ_T:μ_R is derived based on a model or method appropriate to the data, the study design, and the question under consideration. If the confidence interval falls within pre-specified goalposts, the formulations are declared bioequivalent.

Differences in between- or within-subject variation received little attention in the press during this period though references in the work by Haynes (1981) indicate that they were not neglected by the statisticians involved in the bioequivalence debate. Testing procedures for reference and test formulation variability estimates were derived by Pitman (1939) and Morgan (1939) for paired data, such as that arising in cross-over studies. Though potentially of interest, measures of accuracy (i.e. mean bioavailability) seemingly were deemed more important than precision (i.e. variability in bioavailability) as a first step in addressing the evolving question of bioequivalence. Statistical methods had been developed (Marcus et al., 1976) at this time for the closed ordered

testing of hypotheses to ensure fixed experiment-wise error to test such effects in series. Such methodology requires a set of pre-ordered, pre-set hypothesis to be closed under intersection, which seemingly could be applied in a straightforward manner to bioequivalence testing (i.e. test for equivalence of means followed by variances or vice versa). Little additional attention has focussed on application of such a straightforward approach to inference until recently (Barrett et al., 2000).

Illustration of the application of the 80/20 rule and the 75/75 rule follows based on the model described in (2) and based on a study design like the one described in Section 1.1. The data for subjects with data in both periods for AUC (n=45) and Cmax (n=47) in Table 4 were analyzed according to the model in $SAS^{©}$ using simple GLM code as follows:

```
PROC GLM;CLASS SEQUENCE SUBJECT PERIOD REGIMEN;
MODEL LNAUC=SEQUENCE SUBJECT(SEQUENCE) PERIOD REGIMEN;
```

Table 4: Data from a Two Period Cross-over Study Design with Test and Reference Formulations

Subject	Seq	AUC Test	AUC Ref	AUC T:R	Cmax Test	Cmax Ref	Cmax T:R
1	RT	79.34	58.16	1.36	2.827	2.589	1.09
2	TR	150.12	142.29	1.06	5.145	3.216	1.6
3	RT	85.59	69.68	1.23	4.407	2.48	1.78
4	TR	36.95	5	7.39	2.442	0.498	4.9
5	RT	.	121.84	.	.	5.319	.
6	TR	24.53	26.05	0.94	1.442	2.728	0.53
7	TR	22.11	34.64	0.64	2.007	3.309	0.61
8	RT	377.15	208.33	1.81	11.808	9.634	1.23
9	TR	703.83	476.56	1.48	15.133	11.155	1.36
10	RT	14.23	17.22	0.83	1.121	1.855	0.6
11	RT	750.79	1407.9	0.53	6.877	13.615	0.51
12	TR	217.06	176.02	1.23	9.433	8.446	1.12
13	RT	21.27	20.81	1.02	1.055	1.21	0.87
14	TR	40.75	152.4	0.27	1.787	6.231	0.29
15	RT	8.67	.	.	1.084	0.995	1.09
16	TR	52.76	51.57	1.02	3.57	2.445	1.46
17	TR	101.52	23.49	4.32	4.476	1.255	3.57
18	RT	269.4	203.22	1.33	9.618	7.496	1.28
19	TR	37.14	30.54	1.22	2.169	2.613	0.83
20	RT	412.42	386.93	1.07	12.536	16.106	0.78
21	RT	33.89	47.96	0.71	2.129	2.679	0.79
22	TR	143.45	42.69	3.36	5.182	3.031	1.71
23	TR	29.8	29.55	1.01	1.714	1.804	0.95
24	RT	32.59	22.7	1.44	1.853	1.727	1.07
25	TR	63.03	92.94	0.68	3.201	5.645	0.57
26	RT	72.36	44.02	1.64	4.546	3.156	1.44
27	RT	423.05	285.78	1.48	11.167	8.422	1.33
28	TR	.	.	.	0.891	0.531	1.68
29	TR	56.7	21.03	2.7	2.203	1.514	1.46
30	TR	61.18	66.41	0.92	3.617	2.13	1.7
31	RT	20.33	40.6	0.5	1.247	1.9	0.66
32	RT	17.75	19.43	0.91	0.91	1.185	0.77
33	TR	1376.02	1200.28	1.15	27.312	22.068	1.24
34	TR	115.33	135.55	0.85	4.688	7.358	0.64
36	RT	1160.53	1048.6	1.11	17.374	18.976	0.92
37	RT	82.7	107.66	0.77	6.024	5.031	1.2
38	TR	17.34	40.35	0.43	1.072	2.15	0.5
39	RT	928.05	469.73	1.98	14.829	6.962	2.13
40	TR	62.23	64.92	0.96	3.025	3.041	0.99
41	TR	48.99	61.74	0.79	2.706	2.808	0.96
42	TR	53.18	17.51	3.04	3.24	1.702	1.9
43	RT	20.09	14.95	1.34	2.278	0.987	2.31
44	RT	28.47	28.57	1	1.773	1.105	1.6
45	RT	411.72	379.9	1.08	13.81	12.615	1.09
46	TR	.	.	.	1.68	.	.
47	RT	46.88	126.09	0.37	2.339	6.977	0.34
48	TR	98.03	236.17	0.42	3.434	7.378	0.47
49	TR	1070.98	1016.52	1.05	21.517	20.116	1.07
50	RT	106.43	75.43	1.41	4.771	4.925	0.97
R=Reference, T=Test							

Implementation of the general linear model (GLM in $SAS^{©}$ for model (2)) for AUC and Cmax data (only subjects with data from both sessions are included) is shown in Figures 3 and 4 below.

General Linear Models Procedure

Dependent Variable: TPARM LOG(AUC)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	46	142.91505522	3.10684903	15.63	0.0001
Error	43	8.54610668	0.19874667		
Corrected Total	89	151.46116190			

R-Square	C.V.	Root MSE	TPARM Mean
0.943576	9.994867	0.4458101	4.4603908

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SEQUENCE	1	1.09473744	1.09473744	5.51	0.0236
SUBJECT(SEQUENCE)	43	141.54662307	3.29178193	16.56	0.0001
PERIOD	1	0.06212092	0.06212092	0.31	0.5790
REGIMEN	1	0.21157379	0.21157379	1.06	0.3080

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SEQUENCE	1	1.09473744	1.09473744	5.51	0.0236
SUBJECT(SEQUENCE)	43	141.54662307	3.29178193	16.56	0.0001
PERIOD	1	0.05710071	0.05710071	0.29	0.5947
REGIMEN	1	0.21157379	0.21157379	1.06	0.3080

Tests of Hypotheses using the Type I MS for SUBJECT(SEQUENCE) as an error term

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SEQUENCE	1	1.09473744	1.09473744	0.33	0.5672

Parameter	Estimate	T for H0: Parameter=0	Pr > |T|	Std Error of Estimate
T-R	0.09699446	1.03	0.3080	0.09400824

REGIMEN	LOG(AUC) LSMEAN	BACK-TRANSFORMED LSMEAN
T	4.5113	91.04
R	4.4143	82.63

Comparison	Estimate	Standard Error	df	t-value	90% C.I. Lower	90% C.I. Upper
T-R	0.0970	0.0940	43	1.6811	-0.0610	0.2550
T:R	1.1019				0.9408	1.2905

Within subject CV= 46.9%

Figure 3: Analysis of Variance for AUC data presented in Table 4 based on model (2)

General Linear Models Procedure

Dependent Variable: TPARM LOG(CMAX)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	48	73.14134703	1.52377806	9.62	0.0001
Error	45	7.12696055	0.15837690		
Corrected Total	93	80.26830758			

R-Square	C.V.	Root MSE	TPARM Mean
0.911211	30.94203	0.3979660	1.2861661

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SEQUENCE	1	0.29177262	0.29177262	1.84	0.1815
SUBJECT(SEQUENCE)	45	72.77061168	1.61712470	10.21	0.0001
PERIOD	1	0.01827351	0.01827351	0.12	0.7357
REGIMEN	1	0.06068922	0.06068922	0.38	0.5390

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SEQUENCE	1	0.29177262	0.29177262	1.84	0.1815
SUBJECT(SEQUENCE)	45	72.77061168	1.61712470	10.21	0.0001
PERIOD	1	0.01687594	0.01687594	0.11	0.7456
REGIMEN	1	0.06068922	0.06068922	0.38	0.5390

Tests of Hypotheses using the Type I MS for SUBJECT(SEQUENCE) as an error term

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SEQUENCE	1	0.29177262	0.29177262	0.18	0.6730

Parameter	Estimate	T for H0: Parameter=0	Pr > |T|	Std Error of Estimate
T-R	0.05083001	0.62	0.5390	0.08211270

REGIMEN	LOG(CMAX) LSMEAN	BACK-TRANSFORMED LSMEAN
T	1.3128	3.72
R	1.2619	3.53

Comparison	Estimate	Standard Error	df	t-value	90% C.I. Lower	90% C.I. Upper
T-R	0.0508	0.0821	45	1.6794	-0.0871	0.1887
T:R	1.0521				0.9166	1.2077

Within subject CV= 41.4%

Figure 4: Analysis of Variance for Cmax data presented in Table 4 based on model (2)

Note that sequence effects (p=0.5672 and p=0.6730 for AUC and Cmax, respectively) were not significant in this data set suggesting that no factors (such as carryover) are present to confound inference between means (for more discussion on the interpretation of sequence effects in the two-period cross-over see Jones and Kenward, Chapter 2, 1989). The p-values for the comparison of $\hat{\mu}_T$ - $\hat{\mu}_R$ were found to be p=0.3080 (for AUC, $\hat{\sigma}_W$=0.4458) and p=0.5390 (for Cmax, $\hat{\sigma}_W$=0.3980).

As no statistically significant change in AUC or Cmax was identified but sample size was insufficient to demonstrate a twenty percent difference in means with eighty percent power and Type I error rate of five percent, based on a paired testing procedure for means (see Walpole et al., Ch 10.9, 1998), this data set would have failed to demonstrate bioequivalence under the 80/20 rule. Post-hoc power was 66% for AUC and 78% for Cmax.

Note that 78% (thirty-five of forty-five subjects with data in both sessions) and 77% (thirty-six of forty-seven subjects with data in both sessions) of subjects had a test:reference formulation

ratio greater than 0.75 for AUC and Cmax, respectively, indicating that both endpoints were indicative of bioequivalent formulations under the 75/75 rule. Period effects were not statistically significant in this data set (p=0.5497 and p=0.7456 for AUC and Cmax, respectively).

1.3 History of Bioequivalence from 1980 through 1992: Developing Average Bioequivalence

Under the model (2), established in the 1970's, it can be shown (Kullback, 1968) that the statistic

$$I(1:2;\mu) = \frac{(\hat{\mu_T} - \hat{\mu_R})^2}{2(\sigma_W^2)} \sim \chi'^2_1(\lambda) \tag{8}$$

is sufficient to compare the means of the test and reference formulations, based on the assumption that the data are from a bi-variate normal distribution with homogeneous within-subject variance of σ_W^2 between formulations. This statistic was assessed in bioequivalence studies based on the estimate for the non-central F-distribution (Patnaik, 1949; Owen, 1965; Johnson, Kotz, and Balakrishnan, Chapter 30, 1994):

$$\frac{(\hat{\mu_T} - \hat{\mu_R})^2}{(2(\hat{\sigma}_W^2)/n)} \sim F_{1,n-2}(\lambda) \tag{9}$$

where λ is the noncentrality parameter $(\mu_T - \mu_R)^2/(2\sigma_W^2/n)$. It should be noted that under this approach it is assumed that the estimate $(n-2)\hat{\sigma}_W^2/\sigma_W^2$ is assumed to be centrally χ^2-distributed such that the second corresponding non-centrality parameter for the non-central F-distribution (see Johnson, Kotz, and Balakrishnan, Chapter 30, 1994) is assumed to be null for the purposes of bioequivalence testing. Thus equation (9) is usually expressed using the statistic

$$\frac{(\hat{\mu_T} - \hat{\mu_R})}{\sqrt{(2(\hat{\sigma}_W^2)/n)}} \sim t_{n-2}(\sqrt{\lambda}) \tag{10}$$

where t is a non-central t-distribution (Johnson, Kotz, and Balakrishnan, Chapter 31, 1994), and n is the sample size in a randomized, balanced two-period cross-over study. As FDA interrogated the performance of the 80/20 and 75/75 rules and potential alternatives (Colaizzi and Lowenthal, 1986), statistical consideration (Metzler and Huang, 1983) of the ideas inherent in the question of bioequivalence continued using these properties.

Extending the 80/20 rule in consideration of the differences in average response between populations and the approaches developed by Westlake, consideration of methods for decision making in the early 1980's focussed on the use of Bayesian posterior probabilities for the construction of comparisons for μ_T:μ_R. Rodda and Davis (1980) and Mandallaz and Mau (1981) interrogated the decision rules introduced by Westlake (1972-1979) and introduced consideration of this distribution relative to a predetermined goalpost interval of (-δ, δ); bioequivalence was concluded if the posterior probability of falling in this interval was higher than a predetermined probability level, e.g. 0.9. This idea was further developed by Fluehler et al. (1981) who introduced graphical methods to accompany the consideration of the posterior probability and recommended that δ be altered according to the drug under study. Selwyn et al. (1981) and Grieve (1985) developed methods for the Bayesian analysis of the randomised, 2 period cross-over and interrogated the impact of various other factors (carryover, choice of prior distributions) on inference. Reisner and Guttman (1992) developed similar ideas in the engineering field, and Yee (1986) developed a non-Bayesian method for deriving the upper and lower bounds of the probabilities for rejecting bioequivalence.

In summary, these methods used the non-central t- (equation 10) and χ^2-distributions and the model (2) to derive probabilistic statements using Bayes' rule on the posterior probability for the difference of μ_T-μ_R given the data observed to assess the degree of average bioequivalence. It is assumed estimated variances for random between- and within-subject effects from (2) are:

$$\frac{(n-2)(\hat{\sigma}_B^2)}{\sigma_B^2} \sim \chi^2_{n-2} \tag{11}$$

$$\frac{(n-2)(\hat{\sigma}_W^2)}{\sigma_W^2} \sim \chi_{n-2}^2 \tag{12}$$

Between and within-subject variances are assumed to be independent, and fixed effects are normally distributed with mean and nested variance appropriate to the model. Prior distributions must be specified for all model parameters. Nuisance effects (period effects) are integrated out of the log-likelihood function using an appropriate method based on Bayes' function (full details may be found in Selwyn et al., 1981 and Grieve, 1985), and the posterior distribution for $\mu_T - \mu_R$ is derived based on the prior distributions, model, and data based on Bayes' rule (Lindsey, Chapter 8, 1996).

Numerical integration or approximate methods (Selwyn et al., 1981; 1984) were initially proposed for use in implementing these techniques; however, these are known to be subject to various problems (eg. impact of starting values and sensitivity to numerical assumptions, computer intensive), and while the techniques offered substantial benefit in the practical assessment of bioequivalence, their use was not encouraged by Regulators within industry applications. Use of a Bayesian procedure was known to be potentially sensitive to the choice of prior distribution - the classic debate between Bayesians and Frequentists - thus leading to questionable validity in implementation, in this time period, in the eyes of those Regulators charged with protection of public health.

Use of Bayesian inference offers substantial benefit in terms of data exploration (for more discussion, see Breslow, 1990). Recent developments in Markov-Chain-Monte-Carlo based methods known as Gibbs sampling (eg. WINBUGS at http://www.mrc-bsu.cam.ac.uk/bugs/) were developed in the late 1980's and 1990's to easily implement Bayesian methods in a straightforward fashion. Illustration of these methods for normal data models may be found in Gelfand et al. (1990). The data from Table 4 were recently analyzed using such a technique to assess bioequivalence.

Non-informative (i.e. flat) prior distributions for the fixed and random effects were assumed with $N \sim (0,\ 10^6)$ for fixed effects where N is the normal distribution with parameters (mean, variance), and $\Gamma \sim (10^{-3},\ 10^{-3})$ for the inverse random effects (i.e. σ_B^{-2} and σ_W^{-2}) where Γ is the gamma distribution with parameters for (location, scale). Posterior distributions were derived,

plotted, and descriptively summarized using log_e-transformed AUC and Cmax using a Gibbs sampler (100000 iterations were performed) according to the procedure developed by Gelfland et al. (1990) for $\phi = \mu_T$ - μ_R or equivalently on the natural scale $\theta = \exp(\mu_T$ - $\mu_R)$. The predetermined goalpost intervals for the parameters of interest were:

$$\phi \in (-0.2231, 0.2231)$$

$$\theta \in (0.80, 1.25)$$

with posterior probability level $prob(-0.2231 < \phi < 0.2231) = prob(0.80 < \theta < 1.25) \geq 0.90$ denoting an acceptably high level of confidence in bioequivalence. Alternatively, one could consider (Kirkwood and Westlake, 1981; Senn, 2001), the probability of being above a predetermined cut-off denoting an acceptable increase in rate and extent of bioavailability or below a predetermined cut-off denoting an acceptable decrease in rate and extent of bioavailability. Using the above lower and upper bounds for ϕ and θ, the posterior probability level for $prob(\phi < -0.2231) = prob(\theta < 0.80) \leq 0.05$, and the posterior probability level for $prob(\phi > 0.2231) = prob(\theta > 1.25) \leq 0.05$ for the formulations to be bioequivalent.

For AUC, $prob(-0.2231 < \phi < 0.2231) = prob(0.80 < \theta < 1.25) = 0.90$; however, while $prob(\phi > 0.2231) = prob(\theta > 1.25) = 0.09$, $prob(\phi < -0.2231) = prob(\theta < 0.80) = 0.01$. Thus although the posterior probability of falling in the interval was acceptably high, the probability of falling above the upper cut-off of 1.25 could also be construed as unacceptably high (of the order of 0.09). See Table 5 and Figure 5.

Table 5: Statistics and Quantiles for AUC ϕ and θ Posterior Densities based on Data presented in Table 4

Parameter	Mean	SD	5%	50%	95%
ϕ	0.096	0.097	-0.062	0.096	0.255
θ	1.106	0.107	0.940	1.101	1.291

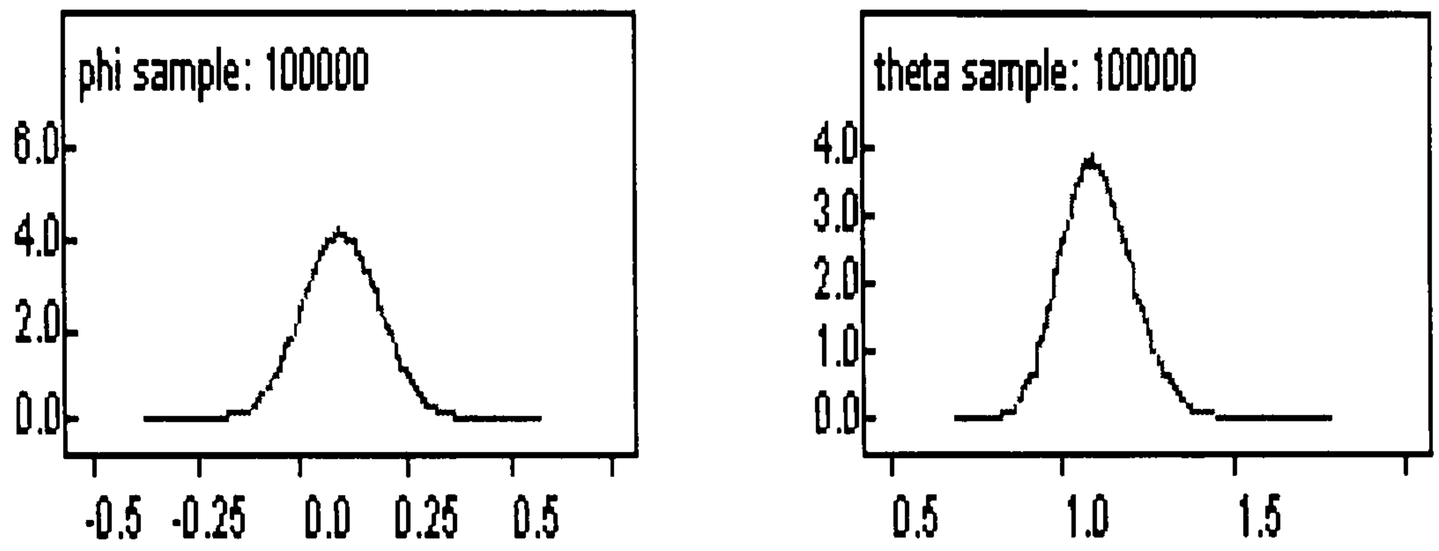


Figure 5: Posterior Densities for AUC ϕ and θ based on Data presented in Table 4

For Cmax, inference is more easily made. The $prob(\text{-}0.2231 < \phi < 0.2231) = prob(0.80 < \theta < 1.25) > 0.90$; and, while $prob(\phi > 0.2231) = prob(\theta > 1.25) < 0.05$, $prob(\phi < \text{-}0.2231) = prob(\theta < 0.80) < 0.05$. Thus the posterior probability of falling within the interval of interest is acceptably high, and the probability of falling above 1.25 or below 0.80 is quite low. See Table 6 and Figure 6.

Table 6: Statistics and Quantiles for Cmax ϕ and θ Posterior Densities based on Data presented in Table 4

Parameter	Mean	SD	5%	50%	95%
ϕ	0.051	0.084	-0.087	0.051	0.189
θ	1.056	0.089	0.917	1.052	1.208

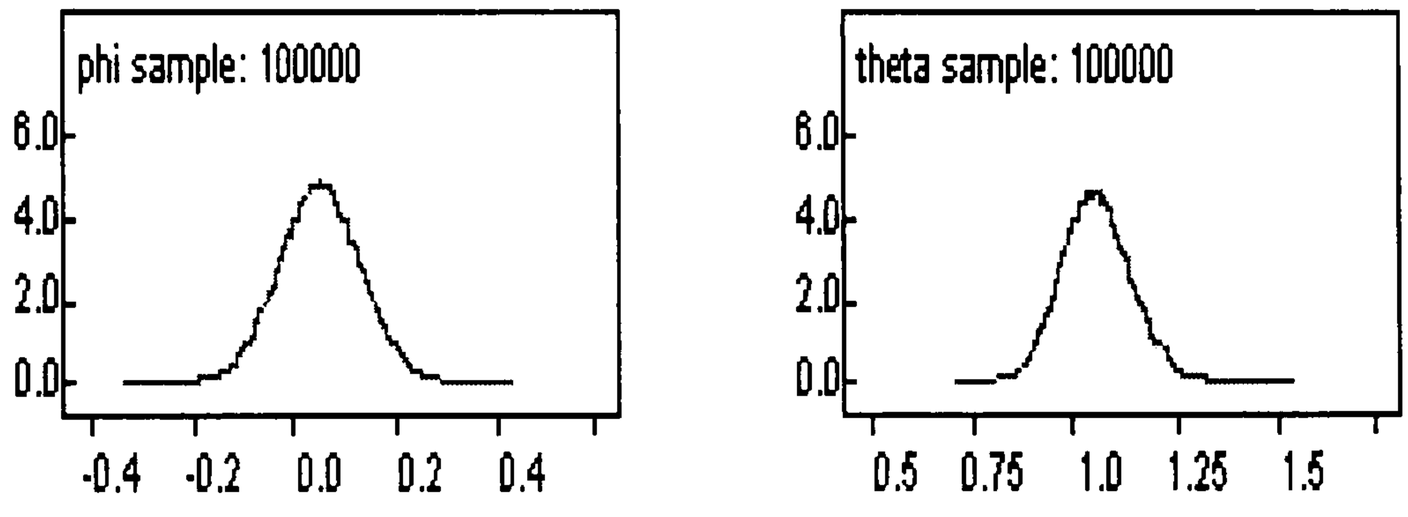


Figure 6: Posterior Densities for Cmax ϕ and θ based on Data presented in Table 4

The response of Food and Drug Administration regulatory authorities to the data generated in this study was interesting. The study was nearly twice the size of most bioequivalence studies (FDA Guidance 1992), but failed to demonstrate bioequivalence under the current standards (to be discussed later in this Chapter) due to unexpectedly large variation. The Bayesian analysis was conducted to characterize the degree of bio-inequivalence and to ask the advice of the regulatory agency on how to proceed. The initial verbal response of the regulators to the data was 'The Biopharm division will not accept a Bayesian analysis as a matter of historical policy (private communication, 1997)'. This is in obvious opposition to the spirit of numerous recent guidances encouraging the use of Bayesian or other alternative techniques where they enhance the understanding of the data (e.g. ICH-E4 Guidance, 1994). This response was amended in the written response (after consulting with the Biostatistical regulatory division) to read, 'Bayesian analysis as described in this submission are not acceptable for bioequivalence assessment (private communication, 1998)'.

Before turning to the statistical procedures which regulatory agencies will accept for bioequivalence assessment, it should be noted that Bayesian analysis naturally facilitates the use of sequential experimentation to assess bioequivalence. Extensions to the Bayesian approach by first conducting a pilot relative biavailability study to estimate within-subject variability in a two-step procedure were discussed in Racine-Poon et al. (1987). A small pilot study (sample size of six subjects) is first conducted under this approach for the purpose of deriving prior beliefs (or distributions). A full size bioequivalence study is then conducted based on this information to assess bioequivalence under predetermined Regulatory standards. Other Bayesian approaches to the assessment of ratios of means are described in Barbieri et al. (2000).

Frequentist implementation of the approaches proposed by Westlake (1972-1979) to the question of bioequivalence were initially assessed by Schuirmann (1981) in the abstract reproduced below:

'In drug-quality control and bioequivalence testing, we may wish to test the 'interval hypothesis'

$$H_O\colon \mu < \theta_1 \text{ or } \mu > \theta_2,$$
$$H_1\colon \theta_1 < \mu < \theta_2$$

at the level α of significance, where θ_1 and θ_2 are known constants, $\theta_1 < \theta_2$. If μ is the mean of a normal distribution with unknown variance σ^2, there does not exist a fixed-sample test with size independent of σ^2. However, we may carry out the usual

size α tests of the hypotheses

$$H_{O1}\colon \mu < \theta_1 \ , \ H_{Or}\colon \mu > \theta_2,$$
$$\text{and } H_{11}\colon \mu \geq \theta_1 \ , \ H_{1r}\colon \mu \leq \theta_2,$$

and reject H_O only if we reject H_{O1} and H_{Or}. The size of this test procedure is always less than the nominal level α. The procedure amounts to rejecting H_O iff the 1-2α confidence interval for μ is completely contained in the interval $[\theta_1, \theta_2]$. We illustrate the power curve for this procedure in several examples and show the actual size depends on the relative length θ_2-θ_1 of the interval, compared to the standard deviation of the estimator of μ. For most cases of practical interest, the actual size is indistinguishable from the nominal size α.' (Schuirmann, 1981)

This procedure was designated the 'two-one sided testing procedure' (TOST), and Schuirmann subsequently refined his work in a publication in 1987, defining the power of the TOST in two-period cross-over designs for the testing of bioequivalence relative to the 80/20 rule for testing bioequivalence.

Blackwelder (1982) and Anderson and Hauck (1983) published similar work. These ideas were further developed in Hauck and Anderson (1984) and Rocke (1984), and general approaches to the question of statistical inference were subsequently summarised under the framework of fiducial probability and inference by O'Quigley and Baudoin (1988). Under this approach to inference, the usual null hypothesis (6) was reformulated to correspond to the structure of testing the question of bioequivalence:

$$H_{01}: \quad \mu_T - \mu_R \leq -\delta \tag{13}$$

$$H_{02}: \quad \mu_T - \mu_R \geq \delta \tag{14}$$

Inference was again based on the use of the non-central t-distribution using a model (typically 2) in a randomised, two-period cross-over design. Summaries of the implementation of such a TOST procedure may be found in Pabst and Jaeger (1990) and Steinijans and Hauschke (1990).

The AUC and Cmax data contained in Table 4 were separately (without multiplicity adjustment in accordance with Hauck et al., 1995) analysed under the model (2) to assess bioequivalence for test relative to reference formulations under the TOST with pre-set limits of $\delta = \ln 1.25$ (corresponding to a twenty percent range on the natural scale). On the natural scale, if the ninety percent confidence interval for $\exp(\hat{\mu_T} - \hat{\mu_R})$ is contained completely within the range 0.80 to 1.25, then bioequivalence is demonstrated. The results were as follows for

AUC and Cmax:

Table 7: Two-One Sided Testing Procedure Results for AUC and Cmax Data in Table 4 from Figures 3 and 4

Endpoint	N	$\exp(\hat{\mu}_T\text{-}\hat{\mu}_R)$	90% Confidence Interval	$\hat{\sigma}_W$
AUC	45	1.10	0.94 - 1.29	0.4458
Cmax	47	1.05	0.92 - 1.21	0.3980

Bioequivalence was not demonstrated in this data set. Cmax equivalence was met under the TOST; however, AUC equivalence was not met as the upper bound fell above the pre-set cutoff value 1.25.

Sample size considerations in the design of bioequivalence studies under the two-one sided testing procedure were developed in Schuirmann (1990), Phillips (1990), and Diletti et al. (1991) based on the bi-variate non-central t-distribution and the model introduced in Section 1.2, (2). As summarized by Phillips (1990), two t-statistics are defined in accordance with the two-one sided hypothesis:

$$T_L = \frac{(\hat{\mu}_T - \hat{\mu}_R - (-\delta))}{\sqrt{(2(\hat{\sigma}^2_W)/n)}} \tag{15}$$

and

$$T_U = \frac{(\hat{\mu}_T - \hat{\mu}_R - \delta)}{\sqrt{(2(\hat{\sigma}^2_W)/n)}} \tag{16}$$

The alternative hypothesis is accepted if T_L and -T_U exceed $t_{1-\alpha,n-2}$ in a balanced, randomized, two-period cross-over design. T_L and T_U have a bi-variate non-central t-distribution (with n-2 degrees of freedom on a balanced design) with non-centrality parameters δ_L and δ_U, such that:

$$\delta_L = \frac{(\mu_T - \mu_R - (-\delta))}{\sqrt{(2(\sigma^2_W)/n)}} \tag{17}$$

and

$$\delta_U = \frac{(\mu_T - \mu_R - \delta)}{\sqrt{(2(\sigma_W^2)/n)}} \tag{18}$$

Under this approach, it can be shown that sample size of n=56 and n=72 subjects are required to demonstrate bioequivalence for Cmax, based on the estimate of within-subject variation for Cmax in Table 7 with 80% and 90% power, respectively, when using a two-period, randomised cross-over. For AUC, n=70 and n=88 subjects are required to demonstrate bioequivalence with 80% and 90% power, respectively under the assumption that no true difference exists between formulations.

Practical considerations in the design and power and sample size of such studies were further developed in Schuirmann (1990). Randomisation to sequence and definition of a wash-out period sufficient to negate potential residual (i.e. carryover) effects from the previous period were established as desirable properties in bioequivalence study design. Sampling scheme in such studies were noted as being very important for proper consideration and definition of Cmax, and period effects were noted as being a 'recurrent phenomenon' in cross-over designs (due to changes in sample storage, environmental conditions, or assay bias between periods - although not significant in the example provided). The use of prospectively designed, properly powered, randomised cross-over designs were established as the norm for bioequivalence assessment.

Regulatory agencies have little direct interest in the power of bioequivalence studies under the TOST (power is typically referred to as 'sponsor's risk' in this context). The Regulator's primary concern is with the confidence level at which bioequivalence can be concluded and with ensuring that the design of such studies ensure an unbiased comparison of formulations. Under Schuirmann's TOST procedure, the confidence level (α) was set at five percent per test for an overall study-wise Type I error rate of up to five percent (FDA Guidance, 1992). The assumptions and requirements underlying bioequivalence assessment in this approach (i.e. the use of general linear models) were stated as:

1. Randomisation of Samples
2. Homogeneity of Variances
3. Additivity (linearity) of the statistical model
4. Independency and normality of residuals

(FDA Guidance, 1992)

FDA thus specified that subjects must be randomised to sequence, and that a general linear model of the form (2) be fit to the log_e-transformed AUC and Cmax for demonstration of bioequivalence in a two-period cross-over design. Between- and within-subject variances were assumed to be homogeneous across formulations, and AUC and Cmax data were assumed to be log-normally distributed. In practical terms, under the 1992 FDA Guidance, equivalence was demonstrated if the 90% confidence interval (calculated using a linear model appropriate to the study design) for $\exp(\mu_T - \mu_R)$ was contained in the interval (0.80-1.25). Different models should be applied if the study design differs from a two-period cross-over design (see Jones and Kenward, 1989; Senn, 1993; Senn, 2002) to construct the confidence intervals.

Firms (also known as sponsors) conducting bioequivalence studies were encouraged to conduct single dose studies at the maximal dose to be marketed in healthy normal subjects ensuring an adequate wash-out period between study periods. AUC and Cmax were designated as the primary endpoints of interest (see Section 1.1) to assess extent and rate of absorption, respectively in the 1992 FDA Guidance.

Log_e-transformed data under this model are assumed to be normally distributed (i.e. log-normally distributed on the natural scale), and goodness of fit can be assessed based upon traditional graphical techniques (plotting studentised residuals versus predicted values and normalized probability plots (Jones and Kenward, Ch 2, 1989) as plotted below. These plots are presented for AUC and Cmax below based on the results of model (2) as presented in Figures 7 and 8.

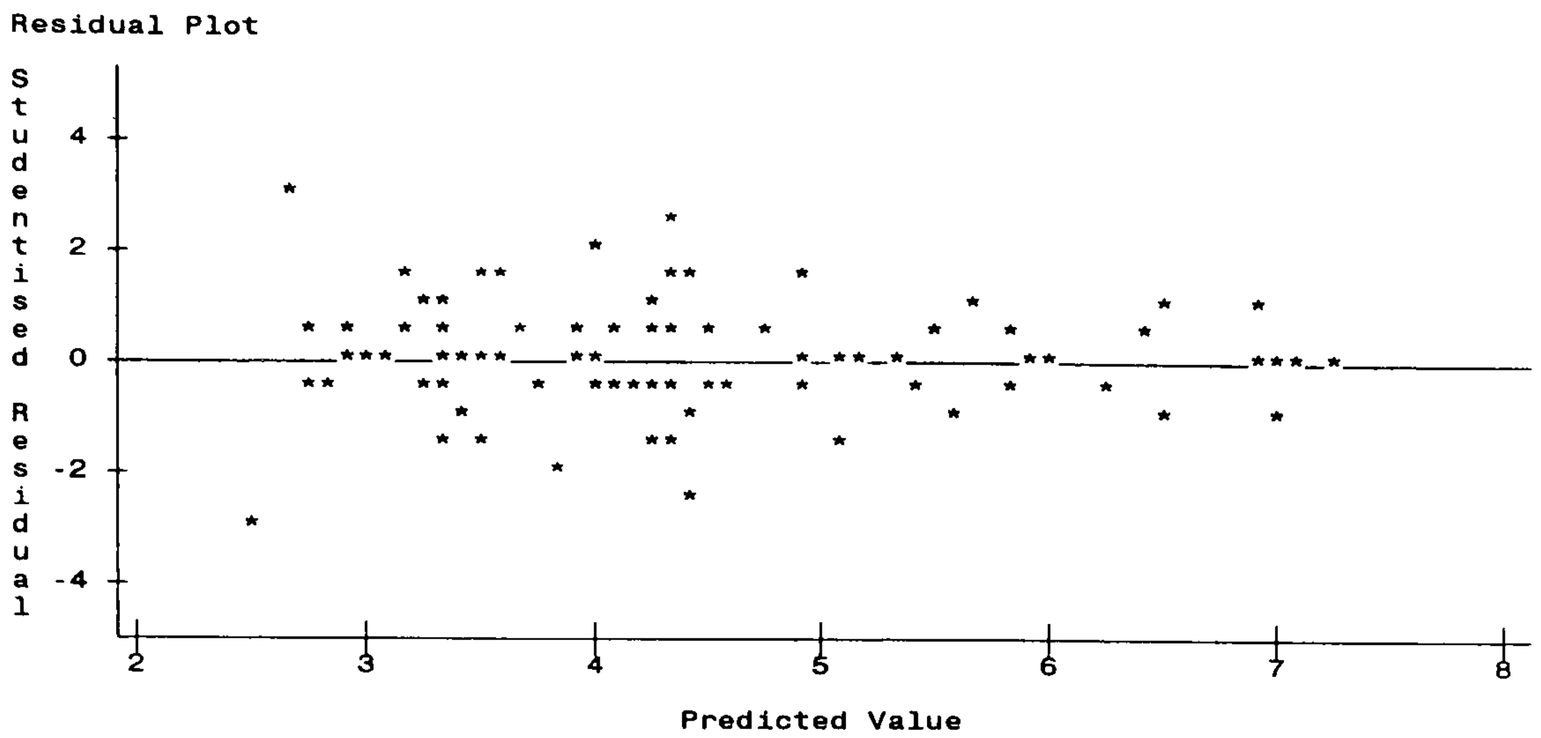


Figure 7: Studentized Residual Plot and Normal Probability Plot for AUC data presented in Table 4 based on model (2)

Residual data for AUC appeared consistent with that of the normal distribution with independent, homogeneous variance. An outlier (subject 004, Table 4) was evident for AUC based on the results of the linear model implemented in GLM (see code earlier in the Chapter). The AUC ratio (Test:Reference) in subject 004 was 7.39; the next most extreme individual high and low ratios were 4.32 and 0.27. When these data are removed and the analysis repeated, the upper end of the confidence interval for AUC comes within the equivalence range (ninety percent confidence interval: 0.91 - 1.22 and point estimate of 1.06).

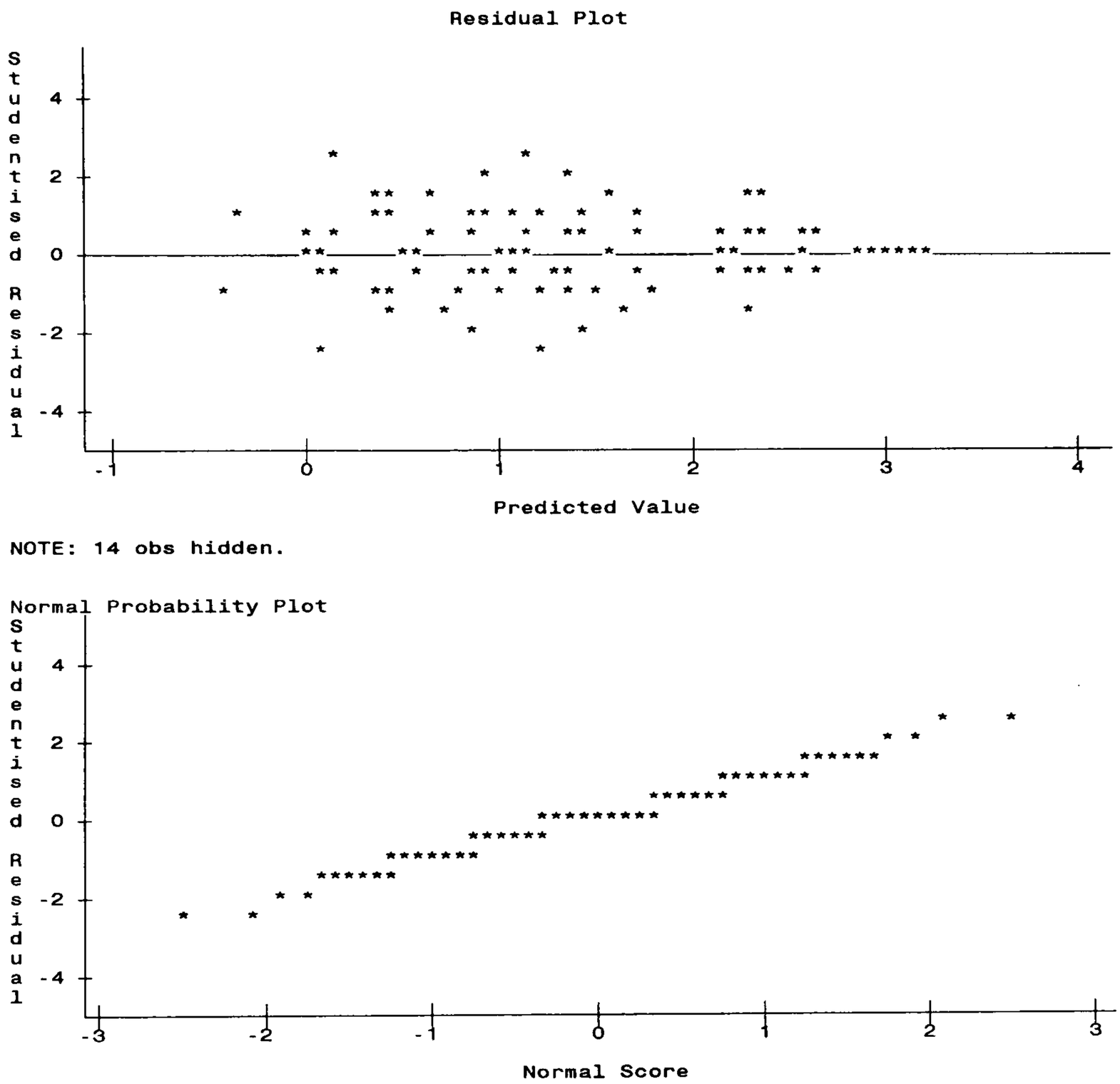


Figure 8: Studentized Residual Plot and Normal Probability Plot for Cmax data presented in Table 4 based on model (2)

Residual data for Cmax appeared consistent with that of the normal distribution with independent, homogeneous variance. No outliers were evident for Cmax based on the results of the linear model implemented in GLM (see code earlier in the Chapter).

Outlier detection (see Jones and Kenward, Ch 2, 1989) was held to be potentially indicative of product failure or subgroup identification (FDA guidance, Chapter V, 1992). It should be noted that statistical detection of an outlier (Lund, 1975) was insufficient reason to exclude a subject's observations; if data were to be excluded 'scientific evidence or explanations' (FDA guidance, Chapter V, 1992) should be supplied. An example of an acceptable reason might be if it could be documented that a subject failed to swallow their medication.

The FDA Guidance (1992) states that, 'Firms are not encouraged to test for normality of data distribution after log transformation, nor should they employ normality of data distribution as a justification for carrying out the statistical analysis on the original scale.' as 'ANOVA models are known to be relatively robust' to deviations from the homogeneity, independency, and normality assumptions (FDA Guidance, 1992). FDA did (earlier in the same guidance) suggest that nonparametric methods be considered (*prior* to ANOVA analysis) in place of the model if data appeared to violate these assumptions; however, in practice, as described in the example below, it appears that only linear-model based methods for log_e-transformed AUC and Cmax data as described in Jones and Kenward (1989) are acceptable without first consulting appropriate regulatory authorities. While estimation procedures more robust to the presence of outliers are known such as M-estimation described by Chi (1994) or Jackknife-estimation by Y. Wang (1999), the acceptability of these procedures in the assessment of bioequivalence is not known.

Nonparametric approaches to inference in a cross-over design had been introduced by Steinijans and Diletti (1983) based upon the Wilcoxon signed rank test (the nonparametric analogue to the t-test) under the assumption that period effects are negligible. Median differences and ninety percent confidence intervals are again constructed on the log_e-transformed scale accounting for each subject as their own control and are exponentiated to provide point estimates and a confidence interval for the ratio of test to reference. In contrast to the normal distributional assumptions underlying the linear model based procedure, this procedure assumes only that data for test and reference products come from independent continuous distributions differing

only in terms of central tendency. Measures of spread are again assumed to be homogeneous between formulations. The downside of this procedure is that the coverage probability is usually higher than that specified (i.e. in excess of ninety percent) resulting in overly wide confidence intervals when the data is truly normally distributed, and power of the procedure to demonstrate bioequivalence is generally lower than the corresponding parametric procedure when the assumptions hold (Hauck et al., 1997).

Nonparametric procedures accounting for potential period effects were subsequently documented by Hauschke et al. (1990) but were subject to the same difficulty in coverage probability. For non-normally distributed data sets however, these intervals may be smaller than those computed based on least squares procedures as shown in Table 8 below.

Table 8: Two-One Sided Testing Procedure Results for AUC and Cmax Data in Table 4 Based on the Non-Parametric Analysis of Hauschke et al. (1990)

Endpoint	N	$\exp(\hat{\mu}_T\text{-}\hat{\mu}_R)$	90% Confidence Interval
AUC	45	1.08	0.95 - 1.22
Cmax	47	1.06	0.92 - 1.21

Bellavance and Tardif (1995) introduced similar techniques for assessment in three-treatment three-period cross-overs.

While Pitman's permutation test can be applied to those data sets where continuous distributions may not be assumed, straightforward exact analysis of binary and categorical data from randomized, two period cross-over designs are described in detail in Jones and Kenward (Chapter 2, 1989). Assessment based upon an additional resampling technique (i.e. the Bootstrap) not requiring the normality assumption was further developed by Efron and Tibshirani (Chapter 25, 1993). Bootstrap-based analysis of bioequivalence studies will be discussed in more detail in Chapter 2-5.

Least squares estimation as recommended in the FDA Guidance (1992 - i.e. ANOVA or GLM based estimation), sometimes referred to as Method-of-Moments based models, are sometimes prone to the estimation of negative variance components in bioavailability designs (Milliken and Johnson, Chapter 22, 1992; Gaffney, 1992), and the interpretation of such mis-estimated, non-negative variance terms is problematic, though not of direct impact to the assessment of average bioequivalence (Vonesh and Chinchilli, Chapter 4, 1997). As noted, estimation approaches using

least squares procedures as in (2) assume that the Huynh-Feldt condition (equality of variances and covariances, Huynh and Feldt, 1970; Hinkelmann and Kempthorne, 1994) holds - an equally troubling assumption in some data sets (as will be shown in Chapter 2). Least squares estimates (Milliken and Johnson, Chapter 22, 1992) for the corresponding between-subject variance are derived as:

$$\hat{\sigma}_B^2 = (\text{MS(Subject(sequence))-MS(Error)})/2$$

resulting in estimates of 1.55 and 0.73 for AUC and Cmax in Table 4, respectively. Within-subject variability estimates ($\hat{\sigma}_W^2$) are 0.20 and 0.16 for AUC and Cmax, respectively.

Cross-over data can also be naturally modelled using restricted maximum likelihood or maximum likelihood techniques (Patterson and Thompson, 1971; Harville, 1977; Laird and Ware, 1982; Brown and Kempton, 1994) as mixed model or repeated measurement data (Jones and Kenward, Chapter 7, 1989; Milliken and Johnson, Chapter 32, 1992; Vonesh and Chinchilli, Chapter 4, 1997). Estimation of the variance-covariance matrix for the mixed effects in such a model are of particular concern (Jones and Kenward, Chapter 7, 1989; Milliken and Johnson, Chapter 22, 1992) and should be specified such that variance components are non-negative. However, proper model building-maximum likelihood based testing procedures (Neyman-Pearson type testing procedures described in Milliken and Johnson, Chapter 1, 1992) to aid in this assessment however are not well established in data sets as small as those usually encountered in bioequivalence studies and should be applied with caution.

Under the assumption that between- and within subject random effects are independent and normally distributed with mean zero, we obtain (after accounting for fixed sequence, period, and formulation effects) that log_e-transformed observations for Reference and Test AUC or Cmax (X_R, X_T) have a bi-variate normal distribution with mean (μ_R, μ_T) and homogeneous variance-covariance matrix with variance $\sigma_B^2 + \sigma_W^2$ and covariance σ_B^2. For a balanced two-period cross-over design, this approach is equivalent to the variance (5) and correlation (4) from model (2), and when data from Table 4 are analysed under this approach, using PROC VARCOMP in $SAS©$ for example, $\hat{\sigma}_B^2$=1.55 and 0.73 and $\hat{\sigma}_W^2$=0.20 and 0.16 for AUC and Cmax, respectively.

Schuirmann's two-one sided testing procedure was adopted as the standard method by European and Canadian regulatory authorities subsequent to finalization of the US FDA guidance in 1992 (Cartwright, 1991; Steinijans, Hauschke, and Schall, 1995). Japan, China, and other

Pacific Rim nations also follow this guidance (with minor changes in study design or decision rules) for the assessment of bioequivalence. Bioequivalence in practice was thus 'harmonised' to assess the difference in means between formulations - this was designated subsequently as 'average' bioequivalence.

Approaches to group-sequential assessment of bioequivalence were discussed using Frequentist procedures by Srinivasan and Langenberg (1986), Kanfer, Geertsema, and Steyn (1988), Snikeris and Tingey (1994), Gould (1995), Whitehead (1996), and Hauck, Preston, and Bois (1997). Such designs are predicated on protection of the Type I error rate (conservation of the probability of incorrectly approving bioequivalence) following numerous looks in the course of the study. Practical implementation of such a group-sequential procedure was discussed in Patterson and Zariffa (2000b) and will be discussed further in Chapter 2.

A method of comparing concentration curves directly through the use of a '*bioequivalence index*' was also introduced by Rescigno (1992) - a method similar to that described by Altman and Bland (1995) - where moments of concentration curves were compared directly to assess association. Similar assessment by the use of moments of the model (2) were developed by Lin (1989, 1992). Under this approach, distance from the complete agreement $\mu_T=\mu_R$ were characterised relative to the regression coefficient ρ. Degree of disagreement of central moments was assessed relative to background variation. The role of ρ in the assessment of bioequivalence will be discussed further in Chapters 2-3.

We now turn to discussion of the use of ABE in the marketplace in the mid-late 1990s and alternative statistical procedures under consideration for the assessment of bioequivalence.

1.4 History of Bioequivalence since 1992: Developing Population and Individual Bioequivalence

The two-one sided testing procedure was easy to implement for nearly any study design and had the benefit of being easy to interpret in practice. In practical terms, the ninety percent confidence interval provides a plausible range of values within which the true difference in means can be expected to fall (Hauck and Anderson, 1986). Use of the procedure quickly became the norm in clinical pharmacology studies of pharmacokinetics as shown in Table 1. Lack of a meaningful pharmacokinetic difference when a drug product was administered with and without

food or with and without a concomitantly administered medication could often be inferred based on the results of such studies under such an approach (Steinijans et al., 1991). Similar techniques could also be used to infer that administration in patients with a concomitant disease state (e.g. hepatic impairment or renal insufficiency) or administration in patients not usually studied in typical pivotal efficacy studies (e.g. a paediatric population) would not result in clinically significant change in AUC or Cmax.

Extent of bioavailability as measured by AUC was judged, under this type of approach, to be a surrogate marker for efficacy in those drugs having been demonstrated to be acceptably efficacious to enter the marketplace. Comparable mean AUC following administration with or without food or a concomitantly administered medication (or in another population) were held to be indicative of efficacy in that condition. Decreases or increases would be used to adjust the dosing strategy for the drug product under study.

Rate of bioavailability as measured by Cmax was held, under this approach, to be a surrogate marker for safety for drugs in the marketplace. Comparable or decreased mean Cmax following administration with or without food or a concomitantly administered medication (or in another population) were held to be indicative of safety. Increases in mean Cmax were potentially suggestive of a less acceptable safety profile for the drug under study.

The range of plausible values as expressed by a confidence interval were used to assess the degree of equivalence or comparability. Confidence level (Type I error) was termed 'consumer' or 'regulator' risk - i.e. the risk of the regulator in making an incorrect decision, allowing market access when the application in fact should not be approved. Though often a prespecified δ was difficult or impossible to define prior to study initiation, inhibiting the ability of study sponsors to adequately ensure adequate power to demonstrate equivalence, power was of less concern when assessing the results of such studies than the confidence level. This gave Regulators an easy standard under which to assess the results of such studies. Choice of whether or not to implement a change in dosing strategy under this approach was often a judgement call on the part of Regulators and was dependent upon their choice of δ.

In contrast, bioequivalence studies were held to a higher standard under the legislation described in Section 1.2. New formulations were not admissible to market unless a successful bioequivalence study demonstrated that they met the regulatory standard under a well-controlled

study using the TOST with predetermined $\delta = \ln 1.25$ (though some nations in Europe allowed a wider standard of $\delta = \ln 1.43$ corresponding to a thirty percent range on the natural scale for Cmax, known to be a more variable endpoint than AUC, see Section 1.1). This *average bioequivalence* approach (so-called as it pertains to the equivalence of the means of the test and reference formulations) has safeguarded public health since its adoption (Barrett et al., 2000).

Equivalence for narrow therapeutic index drugs, those drugs for which a small change in dose or exposure can cause a large alteration in response to treatment, is sometimes regarded as particularly problematic under the average bioequivalence approach (Benet and Goyan, 1995). Examples of such drugs, digoxin and warfarin, (Colaizzi and Lowenthal, 1986) generally exhibit low within-subject variability (i.e. within-subject coefficients of variation less than ten percent.) Under the average bioequivalence approach, it is possible (Phillips, 1990) to demonstrate equivalence of means with prespecified $\delta = \ln 1.25$; however, small average changes in means of statistically significant magnitude are possible. Consider for example a point estimate for mean test to reference formulations of 0.90 with ninety percent confidence interval of 0.85 to 0.97. Such small changes in mean test to reference rate and extent of exposure are potentially clinically meaningful in a proportion of patients (Barrett et al., 2000), and some have advocated (Ansbacher, 1990), special equivalence definitions for narrow therapeutic index products whereby such drugs would be held to a more strict regulatory standard (e.g. equivalence limits corresponding to a ten percent range on the log_e-scale, 0.90 to 1.11).

In contrast, high variability products, defined as those products with within-subject coefficients of variation in excess of 30% (Blume and Midha, 1993), require sample sizes in excess of thirty subjects in order to have eighty to 90% power to demonstrate average bioequivalence in a two period cross-over design (Phillips, 1990). Some have argued (Midha et al., 1997a and 1997b) that small changes in rate and extent of exposure for such products are not clinically meaningful and have advocated allowance of a less strict regulatory standard - e.g. equivalence limits corresponding to a thirty percent equivalence range on the log_e-scale, 0.70 to 1.43, as allowed by European Regulators for Cmax. As an alternative, equivalence limits could be widened based upon the within-subject variability observed in the study (Boddy et al., 1995; Schall and Williams, 1996; Midha et al., 1997a-b) allowing such drug products easier market access.

Structure of within-subject variability in a two-period cross-over thus becomes a question

of concern as it (in combination with the sample size and true mean difference between formulations) determines whether a formulation meets or fails to demonstrate bioequivalence. The structure of this variance term can be explored in several ways.

As is well known, in the general linear models framework, model (2), one can fit the within-subject variance by including a term for subject within sequence by formulation interaction in the model. Alternatively, under a restricted maximum likelihood approach developed in Patterson and Thompson (1971) and in the model notation of Laird and Ware (1980), this approach corresponds to fitting a random-intercept and random-slope model on the assumption that random slope and intercept are normally and independently distributed with null mean and variance of σ_W^2 and σ_B^2, respectively, as follows (Jones and Kenward, 1989; Gaffney, 1992). Let Y_{tj} be the response (log_e-transformed AUC or Cmax) for the j-th subject in the cross-over trial administered formulation t $(t = T, R)$ and

$$Y_{tj} = \mu_t + \nu_j + \varepsilon_{tj} \tag{19}$$

where ν_t and ε_{tj} are independent with mean zero,

$Var(\nu_j) = \sigma_B^2$, the between-subject variance, and

$Var(\varepsilon_{tj}) = \sigma_W^2$, the within-subject variance.

Period and sequence effects would be fitted in the model in practice (see model (2) and Jones and Kenward, Chapter 4, 1989) but are omitted from the description here for the sake of clarity. These two approaches result in the same estimates of variation in balanced data sets with no missing data in two period cross-overs, as discussed in Section 1.3.

Developing this idea further however, assuming that random-effects with mean zero, between-subject variance of (σ_{BT}^2 and σ_{BR}^2), between-subject covariance (σ_{BTR}), and independent within-subject variance (σ_{WT}^2 and σ_{WR}^2) for test (T) and reference (R) formulations are present (though not all moments are estimable in most two-period cross-over designs), the variance of $\hat{\mu_T} - \hat{\mu_R}$ is $(\sigma_{BT}^2 + \sigma_{BR}^2 - 2\sigma_{BTR} + \sigma_{WT}^2 + \sigma_{WR}^2)/n$ (Chinchilli and Esinhart, 1996; Vonesh and Chinchilli, Chapter 4, 1997). Note that under the model (2) developed in Section 1.2, $\sigma_B^2 = \sigma_{BT}^2 = \sigma_{BR}^2 = \sigma_{BTR}$ and $\sigma_W^2 = \sigma_{WT}^2 = \sigma_{WR}^2$ under the Huyhn-Feldt condition (1970). Estimates

for within-subject variation in a two period cross-over study are thus held to be composed of measurement error (not estimable), within-subject variance components (estimable under the Huyhn-Feldt condition), and components associated with between-subject variation (estimable under the Huyhn-Feldt condition).

As an aside we note here that the component of the variance for $\hat{\mu_T} - \hat{\mu_R}$ associated with the variance of differences in between-subject variation estimates $Var(\nu_T - \nu_R) = (\sigma^2_{BT} + \sigma^2_{BR} - 2\sigma_{BTR}) = \sigma^2_D$ is an important consideration in the assessment of what has been termed *individual* bioequivalence and will be discussed later in the Chapter. Here we will only note that under the model (2), this variance is assumed to be null. Operationally, however, it should be noted that between-subject variability is known to be related to the extent of absorption (Hellriegel et al., 1996) complicating assessment of a meaningful difference in between-subject variance (as its magnitude is dependent on the choice of endpoint measuring extent of absorption.)

Average bioequivalence compares the distance between formulations as measured by mean of rate and extent of exposure. Variation under this approach is of secondary interest and generally impacts only the choice of design (when sufficient sample size is considered to provide adequate power) and when assessing the final conclusions of bioequivalence in terms of the distance between means. Increased variation beyond that expected (consider the data presented in Section 1.3) can result in reduced power to demonstrate average bioequivalence. From a sponsor's perspective, therefore, it is preferable to increase sample size to an extent such that if unexplained increases in estimates of variation are observed (e.g. from the presence of an outlier as in the data for AUC in Table 4 or a group of outliers), the sample size is still sufficient to demonstrate bioequivalence in the mean rate and extent of exposure. Outliers are a frequent occurrence in bioequivalence studies and can result from a variety of factors (FDA Guidance, 1992); some may simply be indicative of random variation (i.e. perhaps the volunteers did not heed the protocol requirement to abstain from alcohol intake during the washout period); however, some outliers may be indicative of subgroups in the population who absorb, distribute, metabolize, or eliminate the formulations differently than the general population (for example, see Carter et al., 1993 and Chen et al., 2000b).

The concept of switchability of formulations *for the individual patient* is not addressed by the average bioequivalence criterion (Hwang, et al., 1978). Population means are compared,

and variation between individual subjects (or patients) is factored out of the variation used to assess the distance between population means as described above. Peace (1986), Anderson and Hauck (1990), Hauck and Anderson (1992), and Welleck (1993) introduced the concept of *individual* bioequivalence. Under this approach, the question, 'Can I safely and effectively switch my patient from their current formulation to another?' is addressed using an approach similar to the 75/75 rule discussed in Section 1.2. Under the TIER procedure (Test of Individual Equivalence Ratios) introduced by Anderson and Hauck (1990), a predetermined minimum number of subjects for given sample size and Type I error rate must demonstrate individual ratios for test to reference rate and extent of exposure falling within a predetermined equivalence interval. This approach assumes that period and carryover effects are negligible. The hypothesis that is tested is:

$$H_0 : P_E < MINP \tag{20}$$

versus

$$H_1 : P_E \geq MINP \tag{21}$$

where $MINP$ is the minimum proportion of subjects falling in the predetermined equivalence interval and P_E is the true proportion of equivalent individual ratios. The number of subjects (Y) falling in the equivalence interval is evaluated relative to the null hypothesis using a binomial probability. If the p-value equal to the $Prob$(Number of equivalent subjects $\geq Y$ given $P_E = MINP$ and the sample size in the data set) is less than the pre-set Type I error rate, then bioequivalence under the TIER is demonstrated.

When the TIER is applied to the AUC and Cmax data in Table 4, assuming an equivalence interval for individual test to reference formulations of 0.80 to 1.20, it is observed that sixteen of forty-five subjects and fourteen of forty-seven subjects have individual ratios of test to reference within the equivalence interval for AUC and Cmax, respectively. Assuming a $MINP$ of 0.75, neither AUC nor Cmax demonstrates bioequivalence under the TIER (p-values are approximately unity in both cases.)

TIER based assessment of bioequivalence was discussed in Hwang and Wang (1997). Sensitivity to normal and distributional assumptions was demonstrated; however, as discussed in Section 1.2, these assumptions are not held to be pivotal in the assessment of bioequivalence. Period effects however, (Schuirmann, 1990), are held to be a frequent occurrence in cross-over studies and are a confounding factor in the assessment of individual ratios (Welleck, 1997).

Esinhart and Chinchilli (1994a) developed a method for assessment of an extension of the TIER using tolerance intervals for the ratio of individual responses which accounts for period effects in a two-period cross-over study. A tolerance interval is derived for the ratio of individual ratios using a model accounting for period effects, and should the tolerance interval fall within predetermined acceptance limits, bioequivalence is demonstrated. Assessment under higher order designs was also discussed in Esinhart and Chinchilli (1994a) and is developed in more detail in Chinchilli (1996) and Brown et al. (1997). Sample size determination is described in Esinhart and Chinchilli (1994b). While this method was intuitively attractive, it is evident that sample size requirements for many drug products (those with within-subject coefficient of variation greater than twenty percent) are too great to be addressed in a small, well-controlled two-period clinical pharmacology trial and are still too large to be practically implemented in a higher order design (Esinhart and Chinchilli, 1994b).

Average bioequivalence is a special case of what Hauck and Anderson (1992) have termed *population* bioequivalence. This type of bioequivalence addresses the question, 'Can I safely and effectively start my patient on the currently approved formulation or another?' Differences in variation between formulations should also be considered when determining whether a formulation will be equally effective and safe when administering the commercial formulation of a new drug product relative to that used in clinical trials in Phase III (see Section 1.1). It is not clear in this context whether comparison of within-subject variances or total variances (so termed as the sum of between- and within-subject variance for a given formulation) is the appropriate variance for comparison between formulations, and arguments (Hauck and Anderson, 1994; Grahnen et al., 1984) have been offered for both in this context.

As described in Section 1.2, techniques for comparing within-subject variances in a two period cross-over (under the assumption that between-subject variances across formulations are homogeneous) had been developed by Pitman (1939) and Morgan (1939). Alternatively the

total variances between formulations (between- plus within-subject variance) can be compared using a similar procedure.

Most techniques for assessment of the equality of variances assume that variance components are independent (Brown and Forsythe, 1974; Balakrishnan and Ma, 1989), a condition not met in the correlated data encountered in cross-over trials. Bristol (1991a-b) developed practical maximum likelihood techniques for comparing within-subject variances in this context based on techniques discussed in Mallet (1986). Cornell (1991) derived nonparametric tests of dispersion for the two-period cross-over design. Chow and Liu (Chapter 7, 2000) described similar procedures, and Wang (1997) and Guilbaud (1993, 1999) described similar procedures in subsequent publications. These techniques reduce to different transformations to assess unequal marginal scales in a bi-variate normal population (Kepner and Randles, 1982), and such comparisons were also addressed in work by Bhoj (1979), Ekbohm (1981), McCulloch (1987), and Bauer and Bauer (1994).

When a restricted maximum likelihood model is fitted to the AUC and Cmax data in Table 4, estimates for the variance components are derived as follows (using PROC MIXED in $SAS^{©}$).

Table 9: Estimates of Between- and Within-Subject Variance for AUC and Cmax (Table 4) based on a REML Model

Endpoint	$\hat{\sigma}^2_B$	$\hat{\sigma}^2_{WT}$	$\hat{\sigma}^2_{WR}$
AUC	1.55	0.16	0.24
Cmax	0.73	0.12	0.19

Comparisons of total- or within-subject variance between formulations can be accomplished using such procedures; however, it is known (Zariffa et al., 2000) that variance components are ill-characterized in cross-over studies of the size usually performed. Increasing sample size (Zariffa and Patterson, 2001) can improve the precision of estimated variance components (as we will see in Chapter 5); however, it is unusual for such studies to be performed except in the case of highly variable drug products (Zariffa et al., 2000).

Moreover, while such procedures are theoretically and statistically viable, they are highly dependent (Vonesh and Chinchilli, Chapter 2, 1997) on the choice of estimation procedure. Estimates for between-subject variance can be negative under a method-of-moments based procedure or maximum-likelihood procedure (Bristol, 1991b). Such estimates may be positively

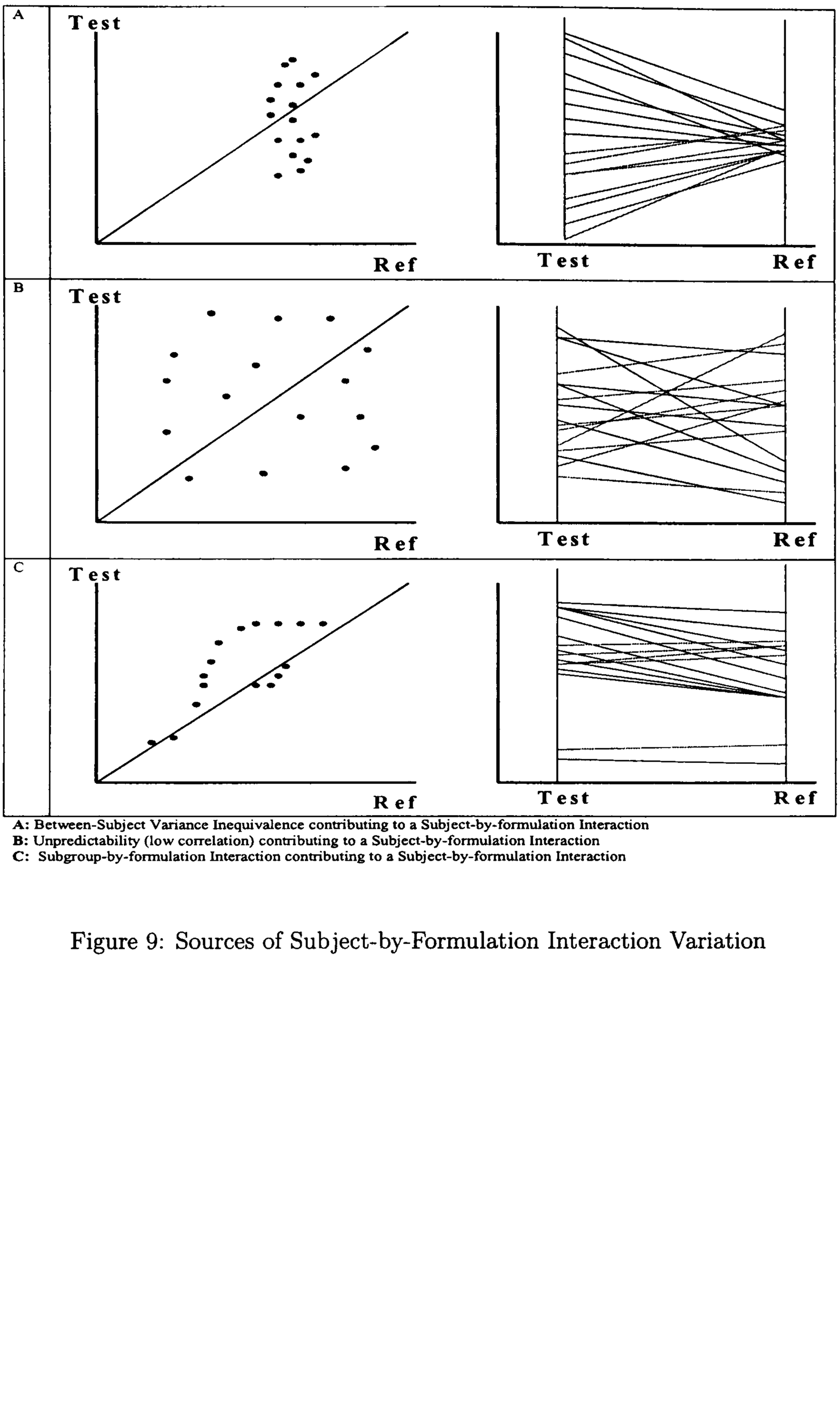
biased (Endrenyi and Tothfalusi, 1999) when using restricted-maximum-likelihood based estimation procedure as would be expected in a procedure constrained in the likelihood to only permit estimates greater than or equal to zero for between-subject variances and correlation constrained to lie in the range [-1, 1] (Patterson and Thompson, 1971; Jones and Kenward, Chapter 7, 1989; Davidian and Giltinan, 1995; Vonesh and Chinchilli, Chapter 4, 1997.) The properties of different models and variance-covariance structures are not well understood in small samples, however, and Chapter 2 will study these features.

Regardless of the poor quality and high dependence of variance component estimation on choice of estimation procedure, such estimates continued to be of interest in the assessment of switchability for bioequivalence (Ekbohm and Melander, 1989.) Under this approach, subject-by-formulation interaction, σ_D^2, quantified as the variance associated with model (2) such that $\sigma_D^2 = (\sigma_{BT}^2 + \sigma_{BR}^2 - 2\sigma_{BTR}) \geq 0$, is termed a measure of individual switchability. Such an estimate is estimable in what is termed a *replicate* design (Gaffney, 1992).

A replicate design is a cross-over study where individual subjects receive a given formulation at least once (Patterson, 1950). Such a design (described in greater detail in Jones and Kenward, Chapter 4, 1989) allows for the estimation of this subject-by-formulation interaction component as it is only partially confounded (Chinchilli and Esinhart, 1996) with within-subject variation for each formulation. Method-of-moment based, maximum likelihood based, or restricted maximum likelihood based estimation procedures (Harville, 1977) can be used to compute the variance components. Within-subject variance estimates can be computed in a straightforward manner based on these procedures (Chinchilli and Esinhart, 1996).

A number of scenarios can give rise to a quantitatively large subject-by-formulation interaction (Hauck et al., 2000). These are presented in the following Figure 9. A classical subject-by-formulation interaction (Ekbohm and Melander, 1989) occurs when subjects experience a more variable response when receiving one formulation relative to the other as illustrated in Figure 9-A. In this example, variation was greater for the test product relative to the reference product. Subject-by-formulation interaction can also occur when unpredictable responses are observed between regimens, as illustrated in Figure 9-B. This is essentially the case when low correlation (where correlation $\rho = \sigma_{BTR}/\sqrt{\sigma_{BT}^2 \sigma_{BR}^2}$) is observed. Subject-by-formulation interact can also be generated from subgroups having differential reactions to drug products as

illustrated in Figure 9-C.



A: Between-Subject Variance Inequivalence contributing to a Subject-by-formulation Interaction
B: Unpredictability (low correlation) contributing to a Subject-by-formulation Interaction
C: Subgroup-by-formulation Interaction contributing to a Subject-by-formulation Interaction

Figure 9: Sources of Subject-by-Formulation Interaction Variation

The use of replicate designs will now be discussed in more detail. Demonstration of average bioequivalence for highly variable drug products requires large numbers of subjects in a standard two-period cross-over (e.g. n =84 for a within-subject coefficient of variation of forty-five percent, n =178 for a within-subject coefficient of variation of seventy percent with ninety percent power). Use of higher-order study designs (e.g. replicate study designs) can make such a seemingly impossibly difficult regulatory hurdle into a probable success.

In a replicate cross-over design, each subject receives each formulation twice as follows. Eligible subjects are randomized to one of two treatment sequences, e.g. TRTR or RTRT (where T denotes the test and R the reference formulations, respectively, see Jones and Kenward, Chapter 4, 1989). Thus, each subject is studied in four periods and receives each formulation twice over the course of the study. Similar to the two period cross-over described above, a washout period adequate to the drug under study (a least five half lives) separates each treatment periods. In each period, the formulation is administered following an overnight fast.

Log_e-transformed AUC and Cmax may be analysed separately using a general linear model with terms for sequence, subject nested within sequence, period, formulation, and the subject(nested within sequence)-by-formulation interaction, commonly termed the subject-by-formulation interaction (Gaffney, 1992). Point estimates for the difference between the test and reference formulations are then calculated. The mean squared errors for subject-by-formulation are used to derive the associated ninety percent confidence intervals. The point and interval estimates are exponentially back-transformed to obtain point and interval estimates of the ratio of test to reference formulations.

In bioequivalence studies, using a replicate design with sequences RTRT and TRTR, the following mixed model for log_e-transformed observations is commonly accepted (Jones and Kenward, 1989). Let Y_{tjk} be the k-th response ($k = 1, 2, \dots$) for the j-th subject in the cross-over trial administered formulation t ($t = T, R$) and

$$Y_{tjk} = \xi_{tj} + \varepsilon_{tjk} = \mu_t + \nu_{tj} + \varepsilon_{tjk} \tag{22}$$

ν_{tj} and ε_{tjk} are independent with mean zero

$$Var(\nu_{tj}) = \sigma^2_{Bt}, \text{ the between-subject variance,}$$

$$Var(\nu_{Tj} - \nu_{Rj}) = \sigma^2_D, \text{ the subject-by-formulation interaction variance,}$$

$$Cov(\nu_{Tj}, \nu_{Rj}) = \rho\sigma_{BT}\sigma_{BR},$$

$$Var(\varepsilon_{tjk}) = \sigma^2_{Wt}, \text{ the within-subject variance,}$$

$$Cov(\varepsilon_{tjk}, \varepsilon_{tjk'}) = 0, \text{ for } k \neq k'.$$

Note that nuisance effects (period and sequence effects) are fit in practice (Jones and Kenward, Chapter 4, 1989) but are omitted from the above description for the sake of clarity. The above model may be fitted using general linear models (corresponding to a method-of-moments approach), maximum likelihood or restricted maximum likelihood based procedures. Differences in estimates between these procedures will be discussed in more detail in Chapter 2.

Under approach to analysis of the replicate design, it can be shown (Vonesh and Chinchilli, Chapter 4, 1997) that the variance for $\hat{\mu_T} - \hat{\mu_R}$ is equal to $(\sigma^2_{BT} + \sigma^2_{BR} - 2\sigma_{BTR} + ((\sigma^2_{WT} + \sigma^2_{WR})/2))/n = (\sigma^2_D + ((\sigma^2_{WT} + \sigma^2_{WR})/2))/n$ in a balanced design with n subjects. Under the assumption that σ^2_D is zero and that within-subject variances are equal, this design is approximately twice as efficient as the two period cross-over (in terms of the sample size required to demonstrate bioequivalence with equal power). It should be noted that the replicate design with sequences RTTR and TRRT is more efficient in those situations where first-order carryover can not be assumed to be negligible (or equal) between formulations (Jones and Kenward, 1989). Note that other four period designs (Jones and Kenward, 1989; Senn and Ezzet, 1999) are as or more efficient than the two-sequence design described above in the presence of first-order carryover.

A replicate design (with seventy five subjects randomized to sequences RTRT and TRTR) was performed with the same formulations used in the study data presented above in Table 4 as that study failed to demonstrate average bioequivalence. Results regarding average bioequivalence were as follows for this second, replicate design, cross-over study.

Table 10: Two-One Sided Testing Procedure Results for the Follow-up Replicate Design Study

Endpoint	N	$\exp(\hat{\mu}_T\text{-}\hat{\mu}_R)$	90% Confidence Interval
AUC	75	0.92	0.84 - 1.01
Cmax	75	0.94	0.86 - 1.03

In this study, average bioequivalence was demonstrated as the ninety percent confidence intervals for AUC and Cmax were contained in the interval 0.80 to 1.25, demonstrating that replicate designs are useful in the assessment of bioequivalence for highly variable drug products. Other findings relating to the magnitude of $\hat{\sigma}_D^2$ and changes in within-subject variances between formulations will be discussed in Chapter 2 where the modelling of replicate designs will be studied in detail.

Sheiner (1992), Schall and Luus (1993), and Schall (1995) introduced an alternative method for individual bioequivalence assessment based on models of dose-response (Sheiner et al., 1989), risk assessment, and different combinations of parameters from the model (22). Under this 'moment-based' approach to bioequivalence assessment, differences in means and variances are combined into one 'aggregate' statistic for the assessment of population and individual bioequivalence (for examples, see Section 1.5). If the upper ninety-five percent bound on the aggregate statistic falls below a preset equivalence margin, bioequivalence was demonstrated. Such a procedure also allows for widening (or narrowing) of the equivalence margin based upon variation observed in the study.

Bootstrap (Schall, 1995) or Bayesian (Sheiner, 1992) based assessment of the quantiles of the composite endpoint were initially proposed; however, estimation procedures for such an aggregate statistic using approximation procedures involving the Cornish-Fisher Expansion (Bickel and Doksum, 1977) and methods for the linear combination of independently χ^2-distributed variables (Huitson, 1955; Fleiss, 1971; Howe, 1974; Harville, 1976; Burdick and Sielken, 1978; Graybill and Wang, 1980; Lu et al., 1988; Ting et al., 1990; Wang, 1990; Burdick and Graybill, 1992) were developed in more detail by Holder (1993) and were published in Holder and Hsuan (1993a-b). Application to the moment-based criterion of most interest (chosen by FDA for implementation, see Section 1.5) was developed in greater detail by Hyslop et al. (1999). An alternative parametric procedure was described by Kimanini and Potvin (1997) and in Quiroz et al. (2002). Practical strategies for population and individual bioequivalence assessment under this approach were developed in Schall and Williams (1996) and will also be discussed in Section 1.5.

Consideration of these ideas led the FDA Biopharmaceutical Science Division (headed by R. Williams) to form a bioequivalence working group in the mid-1990's. This body (composed

of FDA representatives from clinical, scientific, and statistical disciplines) was tasked with determining whether a public health risk under the average bioequivalence approach could exist (Meyer, 1995) and if so to determine a method or methods to evaluate bioequivalence in a manner to protect the public health. A description of the ideas under discussion may be found in Anderson (1993), Hauck and Anderson (1994), Anderson (1995), Anderson and Hauck (1996), Patnaik et al. (1996), Gould (1997), and Chen (1997) but will not be discussed further in this thesis. The conclusion of this debate will be summarised in Section 1.5.

It should be noted that many other statistical approaches were considered during the debate on bioequivalence. Testing procedures for assessing differences in means and variances simultaneously (though not as a composite endpoint) were developed in Bauer and Bauer (1994), Bauer and Keiser (1996), Chen et al. (1996), and Ghosh et al. (1996). Stepwise procedures (testing for equivalence in means between formulations followed by testing for equivalence in variances) were described in Endrenyi and Schulz (1993), Endrenyi (1994), Vuorinen and Turunen (1996), Vuorinen (1997), and Guilbaud (1999), Gould (2000a), and Gould et al. (2000b). Unbiased, optimal tests for bioequivalence assessment were described in Munk (1993), Hsu et al. (1994), Brown et al. (1997), and W. Wang (1999), and multivariate, optimal assessment of bioequivalence (e.g. for AUC and Cmax simultaneously) were described in Berger (1992), Berger and Hsu (1996), Chinchilli and Elswick (1997), and Munk and Pfuger (1999). Testing for differences in profiles were described in Mauger and Chinchilli (2000).

Though statistically valid, under the approach to inference described by Hauck et al. (1995), multivariate procedures were not of direct interest to the bioequivalence debate. The other approaches seem to have little additional benefit in practical bioequivalence assessment relative to those the FDA were considering (Senn, 2000) and thus seem to not have impacted upon the debate.

We now turn to the conclusion of the bioequivalence debate beginning with the draft FDA guidance on population and individual bioequivalence released for public comment in 1997.

1.5 Most Recent Developments: Implementing Population and Individual Bioequivalence

The US Food and Drug Administration's decision following the debate on whether population and individual bioequivalence were needed to protect public health and the approach chosen for assessment were announced in draft guidance released in 1997 (FDA Guidance, 1997) based on the principles discussed in Schall and Williams (1996). Previously discussed approaches (Section 1.4) to moment-based assessment of population and individual bioequivalence were established as described in later paragraphs for studies conducted prior to approval and following approval of new chemical entities. Average bioequivalence was deemed insufficient to protect the public health as it assessed only the difference in formulation means, did not adjust for the variance of narrow therapeutic drug products and highly variable drug products, and did not account for assessment of subject-by-formulation interaction. However, no evidence of therapeutic failure had been established over the five years in which the 1992 FDA guidance had been in effect (Barrett et al., 2000).

Conventional two-period, randomised, well-controlled, cross-over designs were established as the design to be performed in the assessment of population bioequivalence for approval of bioequivalence in formulation changes prior to approval of the new drug product (FDA Guidance, 1997). Two-sequence (RTRT, TRTR), randomised, well controlled, replicate designs (described in Section 1.4) were chosen as the design to be performed in the assessment of individual bioequivalence for approval of new formulations following approval of the new drug product for both generic manufacturers and those manufacturers wishing to make formulation changes following approval. Replicate designs were required for the assessment of individual bioequivalence so that within-subject estimates of variance were estimable along with the subject-by-formulation interaction (FDA Guidance, 1997).

Requirements for adequate washout between study periods were again required to ensure that carryover effects were negligible, and outliers were again deemed to be indicative of either product failure or sub-population-by-formulation interactions (FDA Guidance, 1997). Rate and extent of bioavailability were again measured by Cmax and AUC, respectively.

Overall, the FDA draft Guidance (1997) involved little change in study design for sponsors conducting trials to establish bioequivalence of a new commercial formulation relative to that

used in clinical trials under the population bioequivalence approach to inference though different analyses were recommended for data analysis and decision making. The new draft guidance however required replicate designs for changes in formulations following approval - a more complex design for the majority of drug products. Under this approach to inference, log_e-transformed AUC and Cmax were to be analysed separately using a two stage (mixed effect, restricted maximum likelihood) linear model including terms for sequence, period, and formulation in the model in accordance with model (22) from Section 1.4 for a replicate design. Subject within sequence is specified as a random effect, and a heteroscedastic compound symmetric matrix for between-subject variances is assumed across formulations. Within-subject variability estimates are derived for each formulation. Note that carryover was assumed to be negligible as a feature of the design.

Population bioequivalence was to be assessed using the following aggregate statistic (FDA Guidance, 1997).

$$\frac{(\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2}{max(0.04, \sigma_R^2)} \tag{23}$$

where $\sigma_T^2 = \sigma_{WT}^2 + \sigma_{BT}^2$ and $\sigma_R^2 = \sigma_{WR}^2 + \sigma_{BR}^2$. Note that this aggregate statistic can be constructed using a mixed model from a two period cross-over design (with appropriate modification to model (22)) and does not require the use of a replicate design. This will be developed in Chapter 4.

Individual bioequivalence was to be assessed using the following aggregate statistic (FDA Guidance, 1997).

$$\frac{(\mu_T - \mu_R)^2 + \sigma_D^2 + \sigma_{WT}^2 - \sigma_{WR}^2}{max(0.04, \sigma_{WR}^2)} \tag{24}$$

Because the within-subject variance of each formulation cannot be separately estimated from between-subject variance estimates in most two-period cross-over designs of the form { TR, RT }, a replicate design is generally required. It should be noted that if the Huyhn-Feldt condition is assumed for between-subject variability across formulations, it is sometimes possible to estimate within-subject variability for each formulation using a restricted or maximum likelihood approach to inference. However, under this approach, subject-by-formulation interaction is assumed to be null. The properties of IBE will be developed in Chapter 3.

At least 1500 (2000 samples were recommended in the FDA Guidance, 1997) bootstrap sam-

ples (Efron and Tibshirani, Chapter 25, 1993) preserving the number of subjects in each sequence are derived, and the above mixed model is fit to each bootstrap sample. The appropriate aggregate statistic, either (23) or (24), is derived based on the model estimates for each bootstrap sample; note that the denominator for each bootstrap's aggregate statistic is chosen based on the point estimate from the model estimates of the original data set. The nonparametric percentile method (Efron and Tibshirani, 1993) is then used to calculate an upper ninety five percent bound for the quantity of interest. It was required that the upper ninety-fifth percent bound for the metric of interest fall below predetermined regulatory bounds (1.74 and 2.49 for population and individual bioequivalence, respectively) for both AUC and Cmax for bioequivalence to have been demonstrated. These predetermined bounds were established as follows.

The goalpost for population bioequivalence assessment assumes a total-subject variance for the reference formulation of 0.04 and is set to 1.74 as follows

$$\frac{(log_e(1.25))^2 + (0.02)}{0.04} \tag{25}$$

allowing for a mean difference of twenty percent on the log_e-scale and a variance allowance of 0.05 in the numerator under the procedure proposed by the FDA (FDA Guidance, 1997). If the upper ninety-five percent bound on the FDA metric falls below this value, population bioequivalence is demonstrated for the endpoint under study.

The goalpost for individual bioequivalence assessment assumes a within-subject variance for the reference formulation of 0.04 and is set to 2.49 as follows

$$\frac{(log_e(1.25))^2 + (0.03) + (0.02)}{0.04} \tag{26}$$

allowing for a mean difference of twenty percent and a variance allowance of 0.03 in the numerator for subject-by-formulation interaction and 0.02 for the difference in within-subject variance under the procedure proposed by the FDA (FDA Guidance, 1997). If the upper ninety-five percent bound on the FDA metric falls below this value of 2.49, individual bioequivalence is demonstrated for the endpoint under study.

Responses to release of the US FDA's draft guidance (1997) were plentiful from academia. Scientific flaws of the new procedure for individual bioequivalence were noted (Endrenyi et al.,

1998b) as being:

1. The numerical tradeoff of distance between within-subject variances and the means was strongly asymmetric. Developed in more detail in Endrenyi and Hao (1998c), it was found that a small change in within-subject variances, could allow for a change in means, which would still permit a conclusion of bioequivalence but which would expose a large number of individual patents to risk of therapeutic failure or overexposure to drug.
2. The scaling of the criterion to within-subject variance potentially declares the equivalence of formulations liberally. Again, scaling to variance could allow a proportion of individual patients to exceed safe or therapeutic levels of drug product when switched to a new medication.
3. Computational uncertainty of estimation, both in the models used to assess population and individual bioequivalence [see model (22), and the nonparametric-percentile bootstrap method (Efron and Tibshirani, Chapter 25, 1993)] used to assess inference were noted as being of potential concern when near the predetermined acceptance bound.

Subsequent work describing the properties of the subject-by-formulation interaction in Endrenyi and Tothfalusi (1999) determined that this subject-by-formulation interaction statistic was directly confounded under a restricted-maximum-likelihood based estimation approach, with within-subject variation, though the extent and impact on inference of this bias was not characterised. Bias is not unexpected under such a constrained likelihood based procedure and is clinically meaningful in that between-subject variation is known to be confounded with extent of bioavailability (Hellriegel et al., 1996). Initially, it was thought that method-of-moment based estimates for $\hat{\sigma}_D^2$ (Endrenyi et al., 2000) are unbiased (as we will see in Chapter 5, this is not a safe assumption), but the variance of $\hat{\sigma}_D^2$ is still related to $(\sigma_W^2)^2$ where W denotes within-subject variation pooled across test and reference formulations.

On practical grounds, Endrenyi et al. (1998b) and Endrenyi and Midha (1998d) determined that:

1. Average bioequivalence had not been observed to fail to protect the public health as no objective, adequately demonstrated reports to this effect had been published.
2. It was noted that the available data made public by the FDA from replicate designs for estimates of subject-by-formulation interaction ($\hat{\sigma}_D^2$) were not sufficient to demonstrate a clinical need for the assessment of individual bioequivalence. Furthermore, bioequivalence studies were not conducted in the patient population of concern, and so clinical safety/therapeutic failure could not reasonably be assessed in a non-patient population. As such, comparison of between- and within-subject variances had not been demonstrated to be clinically relevant surrogate markers for therapeutic inefficacy and or unacceptable safety profile.

Other academic responses to the draft FDA guidance (1997) by Senn (1998) noted that it was inappropriate for generic drug products to be held to a stricter standard (i.e. be subject

to assessment of differences in within-subject variance) when innovator drug products were not (i.e. only held to the assessment of differences in total-subject variance between formulations). The new procedures were also noted as being illogical in terms of risk assessment (Senn, 2000). Patients are more at risk when they *start* a new treatment than when they *switch* to a new formulation following ongoing treatment implying that standards for population bioequivalence should be more stringent than individual bioequivalence.

Lindley (1998), following on from ideas originally discussed in Westlake (1986) and Hwang (1996), argued that bioequivalence determination involved making a decision and proposed the use of Bayesian decision theory (Lindley, 1971) in bioequivalence assessment. Lindley discussed two potential decisions: δ_1 to declare bioequivalence, and δ_0 to deny bioequivalence based upon a measure of equivalence, θ. Under this approach, a loss function $L_\theta = u(\delta_0, \theta) - u(\delta_1, \theta)$, predetermined based on agreement between the sponsoring company and regulatory authorities (Lindley and Singpurwalla, 1991), is assessed where $u(\delta, \theta)$ is a utility function measuring the worth of δ when the uncertain value is θ. Expected loss can be derived using a prior distribution for θ and Bayes' rule using available software packages (discussed earlier in this Chapter) and applied to bioequivalence assessment, and structure of the problem can easily be extended to multiple bioequivalence measures (i.e. for AUC and Cmax). Lindley (1998) discussed a straightforward method for choice of prior distributions and describes previous work impacting choice of sample size (Lindley, 1997) under such an approach.

Industry responses on the scientific merits of the FDA draft guidance (1997) were similarly negative and were primarily based on retrospective analysis of existing replicate design data sets. Key findings are summarized below (and will be explored in more detail in Chapters 2-4). preliminary findings results were presented at the American Association of Pharmaceutical Scientists held a joint workshop with the FDA on the topic of bioequivalence from 16-18 March 1998 in Washington D.C. at which a few industry representatives were able to speak to the scientific issues behind the proposal. Of particular note, preliminary analyses of SmithKline Beecham's existing database of previously performed replicate design studies (Zariffa et al., 1998) revealed that:

1. Large differences in means (not permitted under the average bioequivalence criteria) were permitted under the new population and individual approaches to bioequiv-

alence when offset by decreased test variance or scaling to variance of the reference formulation. This was particularly the case for highly variable drug products.
2. Also, substantial subject-by-formulation interaction variation could be masked in the aggregate individual bioequivalence criteria by decreased within-subject test formulation variation relative to within-subject reference formulation variation.
3. Behaviour of the individual bioequivalence statistic when variation for the reference product nears the cutoff (0.04, see (23) and (24)) was inconsistent with logical inference concerning bioequivalence.
4. Results for Cmax were far less consistent between average, population, and individual criteria than AUC suggesting that uniform criteria for rate and extent of bioavailability might not be appropriate.

Additionally, a practical benefit of the new criteria (Zariffa et al., 1998) for sponsors of bioequivalence studies was noted. For highly variable drug products, a substantial decrease in sample size was possible under the new criteria when scaling to the reference formulation's variation. Thus for highly variable products, while a replicate design was required for assessment of bioequivalence, decreased resources would be necessary for sponsors to determine if a formulation was bioequivalent (assuming that subject-by-formulation interaction was negligible).

Subsequent analyses (Patterson et al., 1998) revealed that the effect of scaling to reference product variation was not substantial until coefficients of variation on the order of thirty to forty percent were observed for population and individual bioequivalence. While precision of the estimated variance components was remarkably poor in the existing data sets, it was observed that the magnitude of subject-by-formulation interaction was greater for Cmax than AUC, and the magnitude of subject-by-formulation interaction appeared to increase with increasing magnitude of within-subject variation. The choice of the restricted maximum likelihood estimation procedure in the FDA Guidance (1997) was potentially related to these findings (and will be discussed in Chapter 2).

Other published industry responses to the FDA draft guidance (1997) were similarly negative. Schumaker and Metzler (1998) conducted an analysis of a replicate design study using two formulations of phenytoin (see Section 1.2, phenytoin formulation substitution had been known to be subject to bio-inequivalence in the early 1970's). Notable conclusions of this trial were that, as means between formulation were equivalent and within-subject variation across formulations was homogeneous, no evidence existed for individual bio-inequivalence. This result implied that individual bioequivalence could be assessed using procedures based upon the usual two period cross-over study design, and that the imposition of additional rules for bioequivalence were not necessary. Additionally, as a previously known problematic drug substance was involved, this

implied that the existing FDA Guidance (1992) was sufficient to protect public health.

Responses from European authorities on bioequivalence were similarly negative on the scientific merits of the proposed FDA Guidance (1997). Steinijans and Diletti (1997) noted (see Section 1.4) that comparison of within-subject variance between formulations was possible but that the use of these methods had not properly been considered and that use of the new proposed FDA population and individual bioequivalence criteria had not been justified on clinical grounds. On statistical issues, Steinijans and Diletti (1997) encouraged the consideration of alternative inferential procedures to the use of the bootstrap. Lastly, Steinijans and Diletti (1997) encouraged the FDA to expand its working group and to gain consensus among a wider audience. Similar sentiments were expressed in Godbillon et al. (1996). Later published reports, produced in this period, may be found in Hauschke and Steinijans (2000) and Kimanani et al. (2000a).

In summary, according to the FDA (private communication, 1998), a total of twenty-four individuals provided a total of two-hundred forty separate comments to the FDA draft Guidance (1997) broken down into the following categories (FDA, private communication, 1998):

General:
1. Individual bioequivalence is not justified because the current practice of average bioequivalence has worked well.
2. Subject-by-formulation interactions are unimportant.
3. Individual bioequivalence should not be required for all drugs.
4. Patients should be used in bioequivalence studies rather than healthy volunteers.
5. An individual bioequivalence criterion will not assure interchangeability between two generic products.

Resources:
1. More time to complete a replicate design study.
2. Increased cost of bioequivalence studies.
3. Increase in blood volumes and drug exposure, with possible reduction in availability of subjects.
4. More technical and procedural problems.
5. More subjects have to be recruited because of the high dropout rate.
6. The proposed statistical methods are complicated and would need sophisticated computer software.

Process:
1. The development of the new approaches should be coordinated through the International Conference on Harmonisation.
2. An experimental period is proposed where average bioequivalence is the primary assessment method and the proposed population/individual criteria will be alternatives which may be left to the sponsor's choice.
3. Future studies on approved drugs should use the average bioequivalence approach.
4. The SUPAC-IR (FDA Guidance, 1995) document is not referenced.
5. Lack of harmony between this guidance and the existing food effect guidance.

6. Consensus should be obtained between FDA, Industry, and Academia before broader implementation of the new criteria.

Application:
1. The proposed approach may be suitable only for highly variable drugs, narrow therapeutic index drugs, drugs with long half-life, or special cases where safety and efficacy profiles are greatly affected by the absorption rate of the drug.
2. It is not clear under which certain study designs should be used, e.g. single versus multiple dose.
3. It is not clear that an individual bioequivalence applies during the IND (i.e. pre-approval) phase of drug development.
4. Application of the proposed approach to other clinical pharmacology studies (i.e. drug-drug interactions),
5. Same criterion for metabolites.
6. It would not be necessary to use the individual bioequivalence approach for all drugs, i.e. if the residual variance estimated in a two-period cross-over is adequately small, concerns need not arise about either within-subject variances or subject-by-formulation interaction.

Criterion:
1. A disaggregate criterion might be a better alternative.
2. The mean/variance tradeoff might allow products in the market with substantial mean differences.
3. The proposed individual bioequivalence criterion is asymetric.
4. Justification should be provided for equal weighting of means and variances, as well as for grouping squared differences and differences of squares in the same equation.
5. The interpretation of the aggregate criterion based on transformed values is not straightforward.
6. Why are the criterion not expressed in a more readily interpretable manner?

Methodology:
1. Bootstrap introduces randomness.
2. Bootstrap may be biased if only one is reported.
3. 1500 bootstraps may not be enough.
4. There are many ways to produce random numbers.

Numerous miscellaneous public comments were also received by FDA and will not be discussed in this thesis.

In 1998, FDA subsequently formed what was termed a 'Blue Ribbon Panel' of academic and industry representatives to advise the FDA Working Group on implementation of population and individual bioequivalence in practice. The Pharmaceutical Research and Manufacturers Association (PhRMA) formed a parallel expert panel to assess the issues involved and prepare a joint industry statement on the merits of the proposal. A summary of FDA rejoinders to the concerns of industry, academia, and international regulators may be found in Chen et al. (2000a) and in Williams et al. (2000a).

After considering the public comments on the draft (1997) guidance and after once consulting the Blue Ribbon Panel (October 1998), FDA re-issued two draft guidances on the topic of

bioequivalence in August 1999 (replacing the draft guidance issued in 1997). These two guidances described when to perform a relative bioavailability, population, or individual bioequivalence study (FDA Guidance, 1999a) for drug products in solution, suspensions, aerosols and for topical administration and for the more usual immediate-release and modified-release drug products. General guidance for study design (discussed earlier in this Chapter) were provided. A novel aspect of the guidance was the suggestion that a two-year data collection period for all drugs would be mandated when the guidance was finalised. During this period, all sponsors would be required to perform a replicate design study in order to gain market access, and the sponsoring company would have the option of which criterion to choose to assess bioequivalence (Average or Population bioequivalence for sponsors applying for approval of a new product; Average or Individual bioequivalence for sponsors applying for approval of a new formulation of a product already approved for the market).

FDA acknowledged in the new draft guidance (FDA Guidance, 1999a) that narrow therapeutic index drugs (see discussion in Section 1.4) should be held to a stricter equivalence criteria than the usual twenty-percent range required in the existing FDA Guidance (1992). For these drug products, a ten-percent range on the log_e-scale (corresponding to an equivalence range of 0.90-1.11 on the natural scale) was required.

The second draft guidance from FDA (1999b) described in more detail the study design, model, and approach to statistical inference for average, population, and individual bioequivalence relative to the 1997 draft guidance, but departed from the original approach only in minor respects. Requirements for power and sample size were described in more detail in this draft guidance relative to the original 1997 draft guidance; however, the main departure was in the method of statistical inference.

This draft guidance (1999b) required the use of the Cornish-Fisher expansion (Hyslop et al., 1999; FDA Guidance, 1999b; Hyslop et al., 2000) for the assessment of population and individual bioequivalence based upon estimates derived using a method-of-moments based estimation approach. In contrast, the use of restricted maximum likelihood estimation was required for assessment of average bioequivalence in studies employing a replicate design but was to be used in the assessment of population and individual bioequivalence only in the case of data sets with 'substantial' missing data. Bootstrap based inference was also relegated to the status of

a 'back-up' procedure, to be used only in instances where the Cornish-Fisher expansion and method-of-moments based estimation could provide misleading results. One of the goals of this thesis will be to determine when these different alternatives can and should be employed and to thoroughly characterise the differences in estimation and inference among them and novel alternatives.

Academic responses to the FDA draft (1999a and 1999b) guidances were not as plentiful. Longford (1999) discussed the concept that Phase III pivotal safety/efficacy trials should establish whether treatment effects were of limited variance in the population of interest. If variance was small and only a small proportion of patients could be placed at risk when a novel formulation was introduced, this was not as concerning as those situations where higher levels of variance suggested that a significant proportion of patients would be placed at risk of therapeutic failure. Longford (2000) also introduced an alternative procedure for the assessment of individual bioequivalence based upon a linear combination of independent χ^2 variates where inference could be assessed in small samples using a bootstrap based procedure or in large samples based upon a normal approximation.

Other academic sources (Senn, 2000) held that average bioequivalence should suffice based upon grounds of 'practicality, plausibility, historical adequacy, and purpose' and 'because we have better things to do'. Additionally, Senn (2000) notes that statisticians have 'a bad track record in bioequivalence', that 'the literature is full of ludicrous recommendations from statisticians', that 'Regulatory recommendations (of dubious validity) have been hastily implemented', and that 'Practical realities have been ignored'.

Lastly, other academic authorities (seemingly ignorant of the FDA/AAPS workshops and various data driven publications on the concepts) called for publication of data pertaining to the validity and applicability of the new methods (Colburn and Keefe, 2000).

Innovator and generic industry responses to the newest (1999a-b) draft FDA guidances were however more plentiful. PhRMA's expert panel published its work (2000) and concluded that:

1. The clinical relevance of σ_D^2 and its use as a surrogate marker for switchability could be studied by a targeted clinical pharmacology trial constructed to provide the best evidence of σ_D^2.
2. Trade-offs between parameters, scaling, and the maximum allowable difference (Hauck et al., 1996) could be addressed by the use of an ordered testing procedure.
3. Generic-to-generic switching could be addressed through the use of simulation

studies.

4. To maintain the spirit of global harmonization, it is reasonable to expect that FDA and PhRMA will continue to engage in dialogue with other regulatory agencies and solicit their involvement.

PhRMA's expert panel further recommended that simulation studies be used to assess the use of alternative statistical procedures (Dragalin and Fedorov, 1999a and 1999b; Lin, 1989, 1992, and 2000; Gould, 2000a and 2000b) relative to the FDA draft guidance (1999a and 1999b).

Additional industry responses (Patterson and Zariffa, 1999; Zariffa and Patterson, 1999; Patterson and Zariffa, 2000a; Patterson et al., 2000b; Zariffa et al., 2000) described the practical application of population and individual bioequivalence (to be discussed in more detail in Chapter 2-4) and the behaviour of the proposed criteria based upon actual data and simulation studies. This works concluded that (Zariffa and Patterson, 2001):

Some of the expected features of current proposed population and individual bioequivalence criteria, such as the mean-variance tradeoffs, have been observed in the current database. However, collection of more data in an unsystematic manner will not result in clear answers to the questions of interest. Simulation studies should be utilized to enhance understanding of factors impacting the assessment of bioequivalence and to consider alternative criteria for assessment. Such additional simulation assessment should be undertaken prior to the implementation of a mandatory data collection period. Market access should not be permitted using any new criteria until it is clearly demonstrated that the new criteria offer substantive benefit and no added risk to public health.

It was further recommended that:

There are currently sufficient doubts on both sides of the debate as to the validity of issues raised by the opposing view points. As such, some manner of further study, conducted with scientific rigour, is called for. Given the complex interplay between the many factors at work, it is necessary to clearly outline the goals of proposed further studies to avoid misleading results. Of particular interest is the overall question of added-value, 'Do the proposed criteria reliably address substantial limitations of average bioequivalence and if so, can we be assured they do not in turn introduce additional limitations which could potentially be more serious?' A combination of simulation studies and data collection may be relevant to laying some of the issues to rest.

Simulation Studies:

The authors propose a series of simulation studies be undertaken so as to address fundamental questions regarding the additional value (if any) of the population and individual bioequivalence criteria relative to the existing average bioequivalence criterion. Two key areas of study come to mind: subgroup-by-formulation interaction as well as the appropriate (if any) cut off value for the subject-by-formulation interaction based an appropriate metric. While the recent publication by Hauck et al., (2000) makes an initial attempt to characterise the latter, inherent variability has not been directly included in their study. The ability of traditional statistical

techniques to detect subgroup-by-formulation interaction and that of proposed alternatives PBE and IBE criteria based on 2 period cross-over trials should be studied alongside the current proposed PBE and IBE criteria.

Additional simulation studies related to the sensitivity of the various criteria to single outliers, mean-variance tradeoffs and other comparisons between alternative methodologies proposed (Dragalin and Fedorov, 1999a and 1999b; Lin, 1989, 1992, and 2000; Gould, 2000a and 2000b) are all possible using simulation techniques. Last, distribution of estimated σ_D^2 as a measure of subject-by-formulation interaction under various assumptions would help put into context the relevance of observed values collected over a period of time.

Data Collection:

Since existing data may be of value, one might call for continued efforts in retrospective data analysis. Other sponsors may have similar databases and the FDA has collected replicate design data sets over the past few years. In addition, some of the questions regarding the potential to detect subgroup-by-formulation interactions in 2-period cross-overs can be studied in what must be an extensive FDA repository of such data sets. Once the existing data sets and simulation studies are completed, it may still be required to collect additional replicate design data sets over a fixed period of time.

A minimal set of consideration for any data collection period can be defined as follows. First, and most important, a detailed study protocol should clearly specify the various hypotheses of interest to be addressed in the experimental data collection period. Decisions to be made based upon the data should be pre-specified in the protocol, and criteria for their assessment should be predefined. The protocol should specify the study design and minimum number of studies (and subjects within study) needed to address the hypothesis of interest accounting for both Type I and 2 errors. Given the joint effect between multiple factors, we recommend the minimal sample size in each scenario be set ahead of time. Alternative statistical procedures [Dragalin and Fedorov, 1999a and 1999b; Lin, 1989, 1992, and 2000; Gould, 2000a and 2000b] should be considered.

Finally, the data collection period should be a matter of public record. Data, blinded to compound, should be made available to all interested parties, including relevant key covariates for demography or other factors of interest. The data should be available on an ongoing basis to allow for detailed review by the various stakeholders ahead of any public discussion.

Many of these ideas will be discussed in more detail in Chapter 2-4.

Other industry responses (Kimanani et al., 2000b) criticised the FDA draft guidance on issues of logistics, time, and energy.

International regulatory responses deemed the concepts of population and individual bioequivalence to be un-necessary as average bioequivalence had protected the public health. Representative of this view were the presentations by Ormsby (1999) and Pound (1999) at the FDA/AAPS 1999 Workshop on 'Individual Bioequivalence: Realities and Implementation' co-sponsored by the International Pharmaceutical Federation, Canadian Society for Pharmaceutical Sciences, and the Therapeutic Products Program, Health Canada. Ormsby (1999) noted that until subject-by-formulation interaction had been proven to be indicative of therapeutic failure

and the causes identified, average bioequivalence (which had served to protect the Canadian public since its introduction with over 2500 generic products introduced to the market) would continue to be the standard. Pound (1999) described alteration of the average bioequivalence decision rules with changes in Type I error rate and or acceptance range indicated for narrow therapeutic drug products or those thought to be 'dangerous' in clinical practice.

FDA responses to the questions of interest were plentiful at the FDA/AAPS 1999 Workshop on 'Individual Bioequivalence: Realities and Implementation' but have been discussed previously in this Chapter and will not be re-iterated in this thesis. The chief outcome of the conference was the realisation that little evidence existed to warrant the use of the new bioequivalence methods based on sufficient and adequate safety of patients in the marketplace under average bioequivalence and that subject-by-formulation interaction had not been established as a surrogate marker for therapeutic failure (i.e. there was no 'smoking gun') in an extensive review of replicate design data sets. Population and individual bioequivalence were referred to as a theoretical solution to a theoretical problem.

The FDA Blue Ribbon Panel present at the meeting voted to consider a mandatory data collection period using replicate designs for bioequivalence assessment only for drugs most likely to have a subject-by-formulation interaction, modified release drug products and highly variable drugs, and to allow market access only using the established decision rules of average bioequivalence. This view was subsequently endorsed by the FDA's Advisory Committee of Pharmaceutical Science in September 1999.

Subsequent reports published by the FDA described the rationale behind assessment of subject-by-formulation interaction in the assessment of individual bioequivalence (Hauck et al., 2000). The concept of a large interaction, a $\hat{\sigma}_D^2$ greater than 0.0225, was held to be a conservative measure and potentially indicative of significant subgroup-by-formulation interactions. These concepts will be discussed in more detail in Chapters 3 and 5. Another report (Singh et al., 1999) established population-modeling based procedures for assessing bioequivalence in those drug products where pharmacokinetic measures such as AUC and Cmax cannot be used as surrogate markers for safety and efficacy.

Prior to 2000, FDA still had not established that average bioequivalence had not protected public health; indeed, it would be unusual for a regulatory branch of government to produce

examples of failure of their own procedures to protect the public's interests. Such a situation requires very careful handling by those involved.

FDA did however subsequently produce (Meyer et al., 2000) an example of a data set from a replicate design using two marketed immediate release formulations of methylphenidate (indicated for the treatment of sleep disorders). The innovator version had been admitted to the market following a full clinical development programme; however, the generic version of methylphenidate was admitted following only *in vitro* dissolution testing (under an exception to the average bioequivalence requirements) and had not been held to the average bioequivalence standard. Following reports of therapeutic failure in patients switched to the generic product, a replicate design bioequivalence study was conducted in twenty volunteers. AUC and Cmax of the formulations were bioequivalent under the average bioequivalence approach. However, Cmax showed slightly higher within-subject variance for the test formulation relative to reference was claimed to exhibit a nominally high level of $\hat{\sigma}_D^2$ failing to demonstrate individual bioequivalence under the method proposed by FDA. FDA thus accomplished its goal of producing a 'smoking gun' without having to admit that average bioequivalence had failed to protect public health. Other prospectively performed studies for the assessment of individual bioequivalence were published in Bekersky et al. (1999), Cerutti et al. (1999), Canafax et al. (1999), and Yacobi et al. (2000), but these studies did not identify a difference in formulations.

FDA followed up with this publication in 2000 with the introduction of the 'Biopharmaceutical Classification System' (FDA Guidance, 2000a). Orally administered drug products are categorized based upon in vitro testing into classes I, II, III, or IV. Class I compounds, known as highly soluble and permeable in that they are quick to dissolve when ingested and are absorbed directly into the body quickly, are exempt from the requirements of demonstrating bioequivalence in a clinical study and only must demonstrate that in vitro dissolution profiles for the formulations under study are equivalent. The choice of reference product is of importance in this setting (Spino et al., 2000). The statistical procedures for forming the comparison between products (an application of what is known as the f_2 statistic, FDA Guidance, 2000a) are currently under investigation and will not be discussed further in this thesis. Under the BCS guidance, only Class II, III, and IV drugs are required to demonstrate in vivo bioequivalence before being granted market access. Subsequent discussion at the FDA/AAPS Workshop on Bioavailability

and Bioequivalence (Washington DC, September 2000) revealed that it is likely that European and Japanese Regulators will be developing similar guidances in the future.

FDA guidance (2000b) finalized in October 2000 indicated that the agency would adopt the recommendations of the Pharmaceutical Sciences Advisory Committee (1999). This guidance recommended the use of replicate designs for highly variable and modified release drug products; however, market access was in general to be granted if the study demonstrated average bioequivalence. Alternatively, partial replicate designs might also be utilised (Hyslop and Inglewicz, 2001; Chow et al., 2002). Sponsors conducting the study may use population or individual bioequivalence approaches (FDA Guidance, 2001) to inference if justification is sufficient to meet FDA review. Procedures for review of the data generated by these replicate designs to assess the need and appropriateness for population and individual bioequivalence were under consideration. European guidance (EMEA Draft Guidance, 2001) also continued to utilise average bioequivalence as the standard procedure for bioequivalence assessment but makes no specific recommendation on the validity of population or individual bioequivalence.

Additional reports of therapeutic failure for the product Clozapine, an antipsychotic, were published subsequently (Ereshefsky and Meyer, 2001). Clozapine was granted market access following 'non-standard' bioequivalence studies mandated by FDA under bio-waivers applied for by the manufacturers due to the fact that normal healthy volunteers may not be safely exposed to any dose but the lowest of clozapine. Reports of therapeutic failure followed in the United States where un-controlled switching in-clinic was allowed, resulting in significant costs as this condition requires hospitilisation. FDA subsequently has required the manufacturers of the generic formulations to perform a better bioequivalence study to maintain market access and are preparing a drug specific guidance on the topic of clozapine bioequivalence. Other published reports (Meyer et al., 2001) did not establish a need for individual bioequivalence assessment when considering generic phenytoin products with the innovator product. As discussed earlier in this Chapter, phenytoin switching was known to be problematic in earlier research (see Section 1.2).

Following additional discussion at the 2001 Pharmaceutical Sciences Advisory Committee, the FDA drafted guidance (2002) which removed the potential for using population and individual bioequivalence for market access from their guidance while the criteria were under study.

It is possible that in future the use of these criteria will be re-investigated if FDA determines that there is a need for such based upon observations of the marketplace.

1.6 Summary and Thesis

Bioequivalence studies evolved in the 1960's and 1970's to meet the practical needs of consumers in having access to inexpensive efficacious products and to meet the needs of producers in supplying markets with such products without the extensive costs associated with a full clinical development plan and the delay associated with long clinical studies. On a practical level therefore, their genesis was practical, economic, and driven by legislation to allow market access under strictly regulated conditions. In parallel, scientific advances in drug manufacturing and the science of clinical pharmacology and pharmacokinetics and statistics made it possible to assess differences in mean response between formulations based on small, well-controlled, cross-over studies in normal healthy volunteers.

Therapeutic failures in the 1970's prompted extensive research into the science of bioequivalence. This continued in the 1980's and culminated in the establishment of the techniques for judging formulations bioequivalent based on similarity of mean rate and extent of bioavailability between different formulations of the same drug product. This *average* bioequivalence approach has served to protect the public health since its adoption by the US Food and Drug Administration in 1992 and has quickly spread to all parts of the globe.

However, average bioequivalence compares only the mean rate and extent of bioavailability between formulations and does not compare between- or within-subject variances between formulations. Nor does average bioequivalence assess individual similarity of responses or establish whether different subgroups among the general population will react differently to different formulations. Theoretical solutions to these theoretical problems with the average bioequivalence approach prompted extensive research on the topics of *population* and *individual* bioequivalence in the 1990s. The FDA issued draft guidances for public comment in 1997 and 1999 prompting even more extensive international debate among regulators, academia, and industry.

As of the year 2002, no consensus among regulators, academia, and industry has been established for the use of population and individual bioequivalence. The need for more stringent population and individual bioequivalence has not been demonstrated, and it is known that the

criteria proposed by FDA are actually *less* stringent under certain conditions as we will see in this thesis.

The properties of method-of-moments and restricted maximum likelihood modelling in replicate designs will be explored in Chapter 2, and the application of these techniques to the assessment of average bioequivalence will be considered. Individual and population bioequivalence criteria in replicate cross-over designs will be explored in Chapters 3 and 4, respectively, and retrospective data analysis will be used to characterise the properties and behaviour of the metrics.

Simulation experiments will be conducted in Chapter 5 to address questions arising from the retrospective data analyses in Chapters 2 through 4. Additionally, simulation will be used to explore of a potential phenomenon known as 'bio-creep' - that is the transitivity of individual bioequivalence when multiple generic products enter the market using IBE.

We will then turn to another bioequivalence problem to conclude the thesis; that of comparing rate and extent of exposure between differing ethnic groups as described in ICH-E5 (International Conference on Harmonisation Guidance E5, 1998). The properties of the population bioequivalence metric (FDA Guidance, 2001) and an alternative metric (Kullback, 1968; Dragalin and Fedorov, 1999a) will be characterised in small and large samples from parallel group studies. Inference will be illustrated using data from a recent submission and simulation studies.

Conclusions and areas for further research will be discussed in Chapter 7.

2 Small and Large Sample Properties, Estimation, and Inference for Average Bioequivalence using Replicate Designs

The findings of this chapter were presented at the annual PSI meeting (Patterson and Jones, 2001g), at the annual American Society of Clinical Pharmacology and Therapeutics meeting (Patterson et al., 2000a-b), and in a tutorial at the Drug Information Association meeting (Jones and Patterson, 2002). Aspects of the findings were published in the *Journal of Clinical Pharmacology* (Zariffa and Patterson, 2001), in the *European Journal of Clinical Pharmacology* (Patterson et al., 2001h), in *Pharmaceutical Statistics* (Patterson and Jones, 2002a), and as a GlaxoSmithKline technical report (Patterson and Jones, 2002b).

2.1 Introduction and Goals of Chapter

We now turn to detailed discussion of the use of replicate designs to assess average bioequivalence. Following a brief review, modelling and inferential procedures in average (ABE) bioequivalence will be explored. Simulation will be used to assess hypotheses arising from this exercise (in Chapter 5), and questions remaining to be addressed by additional data collection (and other procedures to be explored in the remainder of this thesis) will be described.

As discussed in Chapter 1, bioequivalence trials (see FDA Guidance, 1992-2002) play an important role in the drug development process. These studies are conducted primarily by pharmaceutical sponsors who have conducted pivotal efficacy trials with a specific formulation of a drug therapy but need or want market access for a more commercially suitable formulation. Also the generic pharmaceutical industry conducts these studies to gain market access for generic formulations of established drug therapies when the patent of the sponsor's formulation expires. The original sponsors themselves may also be required to perform a bioequivalence trial following formulation changes, such as moving site of manufacture.

In these cases, rather than repeat clinical trials to establish the safety and efficacy of the proposed formulations, the pharmacokinetic (PK) characteristics of the plasma-concentration time curve are used to infer that two drug formulations will provide similar therapeutic benefit. The PK is expressed in terms of rate and extent of absorption as characterized by the maximum

observed plasma concentration (Cmax) and the area under the concentration time curve (AUC). In turn, bioequivalence is expressed in terms of the 'similarity' of these two metrics between the two formulations. For approximately the past 10 years, international regulatory agencies in North America, Europe, and Asia have used a criterion of average bioequivalence (ABE) with regulatory limits of 20% (FDA Guidance, 1992-2002) except in instances where dissolution profiles suffice (FDA Guidance, 2000a).

This criterion focuses on the average PK metrics of the two formulations being studied. The framework for statistical inference is based on exact 90% confidence intervals for the difference in formulation means. This ABE criterion has been used for both instances of pre-market approval and post-market formulation changes described above.

To review, average bioequivalence (ABE; FDA Guidance, 1992) has traditionally been used as the standard for market access with regulatory limits of twenty percent. This approach focuses on the average PK metrics of the two formulations being studied. The framework for statistical inference is based on ninety percent confidence intervals for the mean differences and is currently being used for both instances of pre-market approval and post-market formulation changes described above. The two one-sided hypotheses of interest (Schuirmann, 1987) are:

$$H_{01} : \mu_T - \mu_R \leq -\ln 1.25$$

$$H_{02} : \mu_T - \mu_R \geq \ln 1.25$$

where μ_T and μ_R represent mean of the Test and Reference formulations, respectively, and the limit of $\ln 1.25$ is chosen to represent a twenty percent range on the log_e-scale. PK data such as AUC and Cmax are typically held to be *log*-normally distributed (Westlake, 1986) and are treated as normally distributed following appropriate transformation (usually *log* to base e transformation). Inference is based on the use of the non-central, bi-variate t-distribution using a model appropriate to the randomised, cross-over design. AUC and Cmax are analyzed separately, and each one-sided test is performed at the five percent level. No adjustments for multiplicity are made (Hauck et al., 1995).

In practical terms, a ninety percent confidence interval is constructed using the $\hat{\mu}_T - \hat{\mu}_R$

and estimates of variation for $\mu_T - \mu_R$. If the confidence intervals for both AUC and Cmax fall within the range $-\ln 1.25$ to $\ln 1.25$, then average bioequivalence is demonstrated. More commonly, these differences and confidence intervals are exponentiated and assessed relative to the interval 0.80 to 1.25.

Under this approach, eligible subjects (typically normal healthy volunteers) are randomized to one of two treatment sequences, test followed by reference (TR) or reference followed by test (RT). A washout period adequate to the drug under study (a least 5 half lives) separates the treatment periods. In each period, the formulation is administered following an overnight fast.

The models used for testing this ABE procedure are the subject of further research in this Chapter.

2.2 Power and Sample Size

Type I error in average bioequivalence testing is the probability of incorrectly concluding, based on the results of a study, that two formulations are bioequivalent when in fact they are not. This probability is sometimes referred to as 'confidence', 'alpha', or 'regulatory risk'. Each one-sided test in a bioequivalence study is constrained, under regulatory guidance (FDA Guidance 1992-2002) to 5% per test corresponding to a 90% confidence interval.

Power in a bioequivalence study is the sponsor's likelihood of correctly demonstrating bioequivalence when it, in fact, exists (Owen, 1965; Phillips, 1990). Sponsors are interested in maximizing their chances of success, subject to resource constraints. Sample size is chosen based on the Type I error rate (fixed at 5% per test) and the equivalence criteria (as described above), intra-subject variation, and Power (usually fixed at 90%). Sample size requirements increase dramatically when intra-subject coefficients of variation increase beyond 30%, see Table 11. A comprehensive description of these inter-relationships may be found in Senn (1997), and we will comment on some aspects in the following.

An additional complicating factor may also play into the design of such studies. Typically it is reasonable to assume that formulations exhibiting identical dissolution profiles will have true ratios of bioavailability (BA) for the test to reference products equal to unity (i.e. on the log_e-scale, it is assumed that $\mu_T - \mu_R = 0$). However, some drug products may not meet this expectation (the true ratio of bioavailability will be expected to deviate from unity by a small

amount). This requires larger sample sizes to compensate and maintain power, as illustrated in Table 11. When determining sample size, best estimates of intra-subject variability and true ratios are used. Deviations from these assumptions will cause variations in the power of the study's testing procedures.

Properly applied, alternative study designs can be used to reduce the number of subjects required to a more manageable level and ensure conclusive results in a study when it is designed under conditions of uncertainty with regard to assumptions. Some alternatives are the replicate cross-over design and group-sequential designs.

In a replicate cross-over design, each subject receives each formulation twice as follows. Eligible subjects are randomized to one of two treatment sequences, e.g. TRTR or RTRT. Thus, each subject is studied in four periods and receives each formulation twice over the course of the study. Similar to the two period cross-over described above, a washout period adequate to the drug under study (a least 5 half lives) separates each of the four treatment periods. Plasma concentration-time profiles are obtained after each administration, and non-compartmental methods were used to derive summary measures AUC and Cmax. The regulatory decision rule for demonstrating bioequivalence under this design is the same as the two-period cross-over presently, though alternatives (FDA Draft Guidance 1997, 1999a, 1999b; FDA Guidance, 2000b) have been considered. These alternatives have been the subject of international debate for some time (see Chapter 1) and will be discussed in Chapters 3-5.

The number of subjects required to demonstrate average bioequivalence can be reduced by up to 50% using a replicate design, see Table 11, relative to the usual two-period cross-over design. Note that the overall number of doses studied (and blood sampling), however, remains the similar to a two period cross-over and that the study will be of longer duration.

Table 11: Sample Sizes Providing Ninety Percent Power in Bioequivalence Studies for Two Period and Replicate Designs

$CV_W\%$	% Deviation in True BA Ratio	Two-Period Cross-over#	Replicate Design#	Two-Period Cross-over&	Replicate Design&
30	0	40	22	50	26
	5	54	28	60	34
	10	112	56	124	70
45	0	84	44	90	48
	5	112	56	120	64
	10	230	116	244	132
60	0	140	70	146	76
	5	184	92	194	102
	10	384	192	404	210
75	0	200	100	206	106
	5	264	134	276	144
	10	554	278	574	298
#Assumes subject by formulation interaction (see Chapter 1) is negligible					
&Assumes subject by formulation interaction is non-negligible (Hauck et al., 2000)					

Computer code to derive power in two-period cross-over designs is available in a variety of commercially available software packages; however, customised code is required for a replicate design. *SAS*© code for power calculations are provided below:

```
%macro bepower;
DATA indata; n=; vard=; ratio=; cv=; run;
data b; set indata; a=0.05;
s=sqrt(vard+((log((cv/100)**2+1)))); n2=n-2; run;
data outcome; set b; t1=tinv(1-a,n-2); t2=-1*t1;
tau1=(sqrt(n))*((log(ratio)-log(0.8))/s);
tau2=(sqrt(n))*((log(ratio)-log(1.25))/s);
r=(sqrt(n-2))*((tau1-tau2)/(t1-t2));
prob1=probt(t1,r,tau1);
prob2=probt(t2,r,tau2);
answer=prob2-prob1;
power=answer*100; run;
%mend bepower;
```

This code is easily adapted to address alternative study designs with modification; however, power is not a central issue of this research and will not be investigated further here.

Estimated formulation means have error associated with measurement and sampling error in cross-over designs. Replication of measurement within each subject reduces sampling error by a factor equivalent to the number of replications. For example, in a standard two-period cross-over design, variance of an individual's mean response on i =T (Test formulation) or R (Reference formulation) is $\sigma^2_{Bi} + \sigma^2_{Wi}$ where σ^2_{Bi} is the inter-subject variance and σ^2_{Wi} is the intra-subject (i.e. sampling error) variance. In a replicate design, variance of an individual's mean response is $\sigma^2_{Bi}+(\sigma^2_{Wi}/2)$. Therefore, where high intra-subject variability is of concern, the replicate design will provide more precise estimates of the true individual response, see Figure 10. For a low variability product, replication does not improve precision dramatically; however, for a high variability product, replication constrains the range over which an individual's mean response may vary. Such measurement is also more accurate as replicate measurement and the derivation of corresponding means converges to the true (and unknown) mean under the central-limit-theorem with increasing replication (Walpole et al., 1998). Such measurement may thus allow for better scrutiny of outliers (Williams et al., 2000b), but as comparison of formulation means is of direct concern in the success of average bioequivalence studies, the desirability of such improvement in accuracy and precision is immediately apparent as a practical matter.

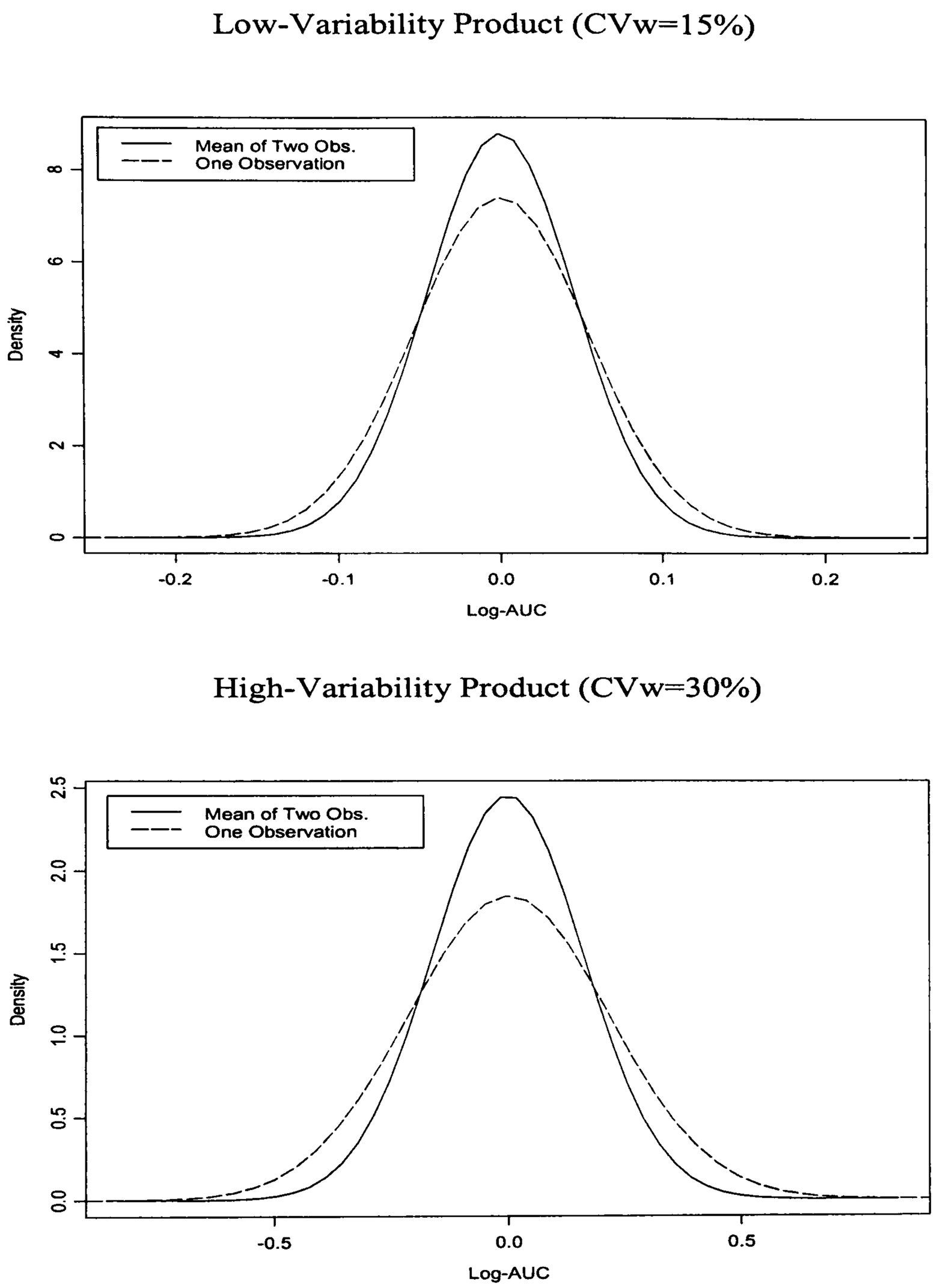


Figure 10: Improvement in Precision due to Intra-Subject Replication for a Low Variability and High Variability Product; $CV_W = \sqrt{\exp(\sigma_W^2) - 1}$ =Intra-Subject Coefficient of Variation

2.3 Discussion of Extension to Group-Sequential Designs

Group sequential designs offer the potential for additional resource savings in bioequivalence designs (Gould, 1995; Hauck et al., 1997). A group sequential design consists of one or more interim analyses (see Figure 11), at which point the sponsor can decide to stop the trial with concrete evidence of success or failure or to carry on. Well known in the statistics community (Peace, 1992), such designs are easy and straightforward to implement in practice in this setting.

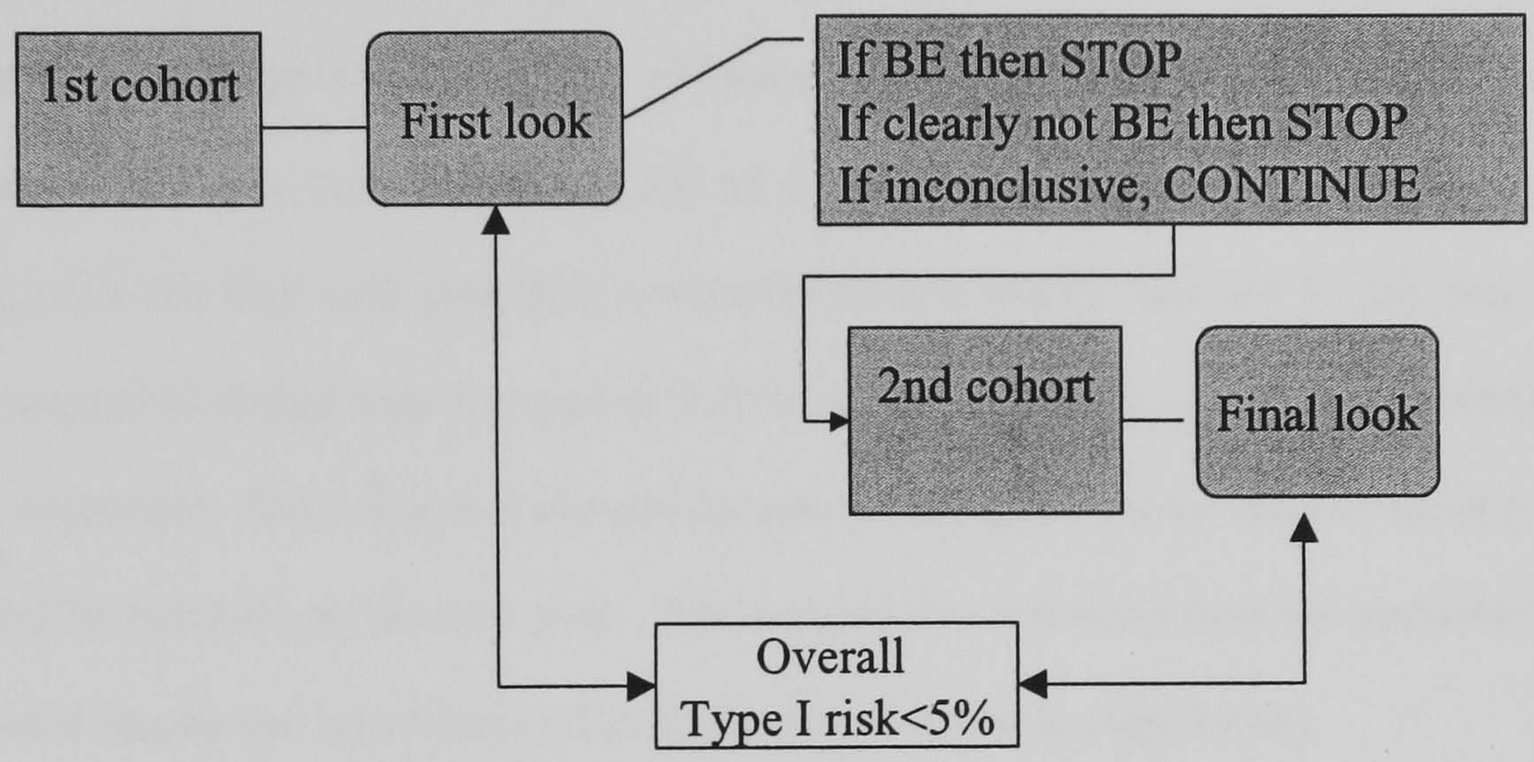


Figure 11: The Concept of Group-Sequential Designs

A group sequential design approach could be used in cases where there is some uncertainty about estimates of variability. That is, based on previous data there is a fairly wide range of estimates, such that choosing a lower estimate might result in an under-powered study and choosing a higher estimate might result in an over-powered study, which in either case is a waste of resources. As such, the group sequential design allows one to conduct an interim look with a sample size that provides reasonable power based on a lower (or optimistic) estimate of variability and the final sample size based on a higher (or less optimistic) estimate of variability. Similarly, if uncertainty in the true ratio of bioavailability is of concern, an interim look might be planned based upon sample size required to provide bioequivalence based on the optimistic estimate, with the final look providing conclusive results should this not be the case. Lastly, a group-sequential design may be applied if it is undesirable to complete a large study due to resource constraint. Some choice of samples for interim analysis may be chosen (based on clinical feasibility) to facilitate an interim look. The probability of success may be quantified at that stage, and if results are inconclusive, the study can continue to completion.

The two aspects of a group sequential design that help determine the probability of stopping early are the alpha-spending function to control the overall Type I error rate of the study and the decision rule(s) for stopping at an interim analysis. There are many Type I error spending functions and decision rules to chose from, but only those relevant to two-stage group sequential design for a bioequivalence trial will be discussed in this paper.

Type I error rate (usually set by regulators at 5% per test for bioequivalence studies) is defined as the probability of a false-positive outcome, or in the case of bioequivalence trails, declaring two formulations are bioequivalent when they are not in truth. Unlike a fixed sample size trial where there is only one analysis, a group sequential trial may have multiple analyses. When data from a fixed sample size trial are analyzed repeatedly during the trial, the overall Type I error becomes inflated if each look is conducted at the same test level. For example, if two bioequivalence test procedures are is conducted (each at the usual 5% level), the overall Type I error rate, the probability of a false positive on the first or second test, is 8% (instead of 5%); if three are conducted, the overall rate is 11%; and so on (Wetherill and Glazebrook, 1986).

As such, to control the overall Type I error rate of the study, the Type I error rate at each

analysis must be some value less than the desired overall Type I error rate. In a two-stage group sequential bioequivalence trial, the Type I error is typically divided equally between the two analyses. A simple, but conservative, method is the Bonferroni adjustment, which results in an error rate of 2.5% (i.e. 95% CI) at each look, but an overall error rate less than 5%. Another alternative suggested by Pocock (1977), is to set the error rate at the two analyses at 2.94% (i.e. approximately 94% CI) at each look, resulting in an overall error rate of approximately 5%. Stopping rules should be defined when the study is designed, and implications of their choice should be considered for impact on sample size (for more discussion see Gould, 1995).

The decision rule for stopping early (at the first look) should contain both a rule for stopping early when bioequivalence is clearly demonstrated and a rule for when bioequivalence is not expected to be demonstrated, see Table 12.

Table 12: Practical Stopping Rules for An Interim Look in a Group-sequential Bioequivalence Study

Outcome	Action
Test regimen is BE i.e. the 95% CIs are contained in (0.80-1.25)	Success. Stop the study and accept BE.
At least 1 Point Estimate is outside the range (0.80-1.25)	Futility. Stop the study and reject BE.
Point estimates are in the range but CIs are not	Inconclusive. Continue the study.

2.4 Design Considerations and Examples

An algorithm for designing average bioequivalence trials is described below.

1. Calculate power for available sample size (ie. number of beds and other clinical resources) for the bioequivalence trial for a standard two-period cross-over design.

2. Consider available resources relative to desired probability of success (power) and re-evaluate choice of sample size.

2a. For products with low-moderate intra-subject variation where adequate resources are available, use a standard two-period cross-over design.

2b. For highly-variable products, where sample size exceeds available resources,

consider a replicate design and re-assess sample size. If resources are adequate, use the replicate design.

3. For situations where resources are still too limited to achieve desired power, or in situations where one is uncertain of the magnitude of intra-subject variation (or other assumptions), consider a group-sequential design.

To illustrate, in a recent site transfer for a highly variable drug product, it was required that ABE be demonstrated at each of three dose levels (Patterson et al., 2000b). For this drug substance, intra-subject variance appeared to fluctuate with dose (possibly increasing as dose was increased.) At lower doses, based on previous experience with relative bioavailability studies, the true ratio in test to reference bioavailability was expected to deviate by up to 5% from unity.

Based on these factors, it was decided to implement a replicate design in these studies. A 10% dropout rate was assumed and accounted for in sample size calculations. Eighty normal healthy volunteer subjects were recruited in order to complete at least seventy-two subjects in a replicate design cross-over in studies I and II.

As sample size was still prohibitively large relative to available resources for the third study, Study III was conducted using a group-sequential, replicate design, see Figure 12. The conservative Bonferroni adjustment to Type I error was employed at each look (Type I error rate of 2.5% per one-sided test corresponding operationally to a 95% confidence interval.) The bioequivalence decision rule to be followed the first look in this study is summarized in Table 12.

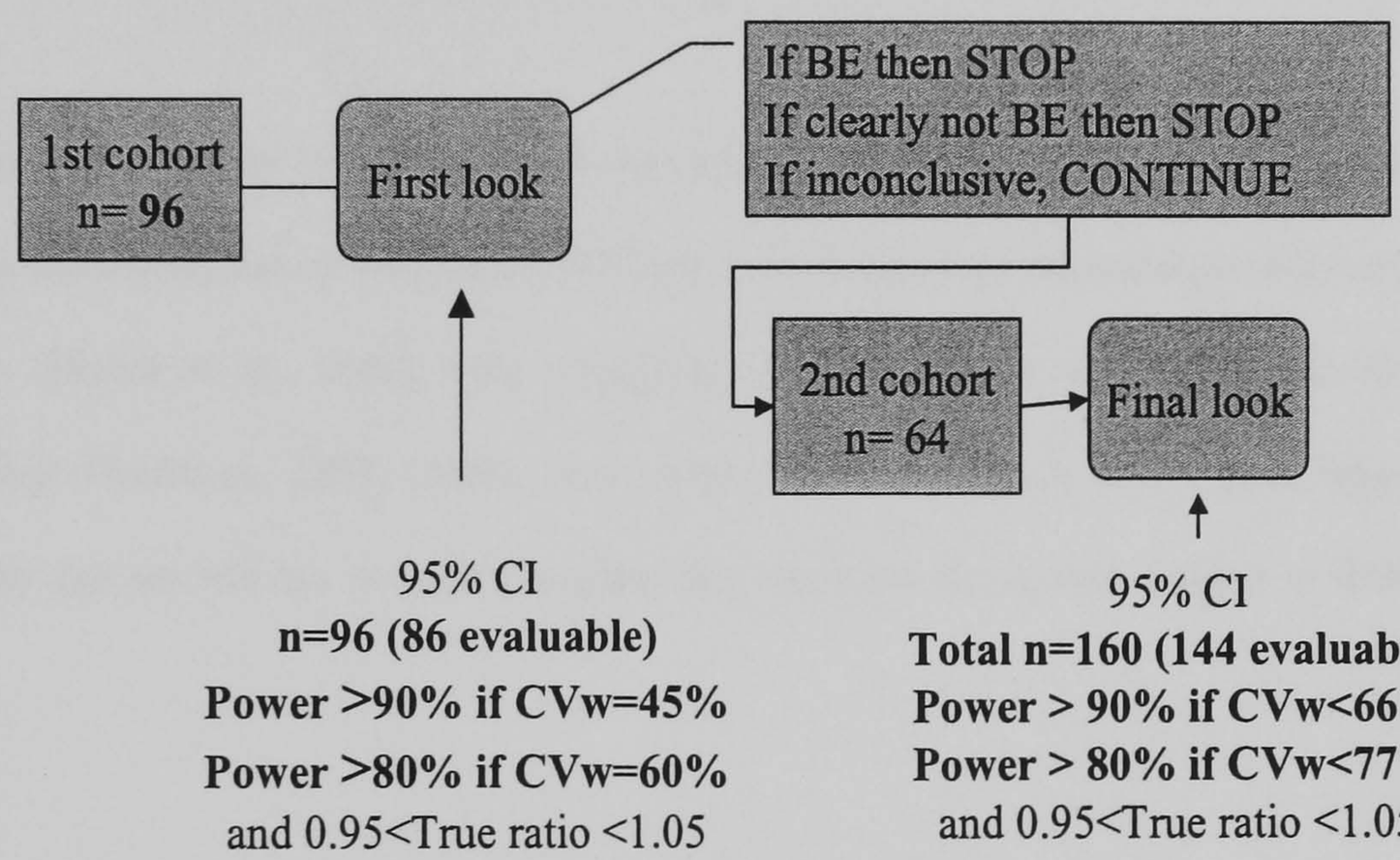


Figure 12: Application of a Group-Sequential Design for Study III

Statistical analyses were conducted using $SAS^{©}$ version 6.12 on a COMPAQ Deskpro computer in accordance with the models for a replicate design described later in this Chapter.

Average bioequivalence was demonstrated conclusively in all three studies (see Table 13). Study III demonstrated bioequivalence at the first interim look. Had the criteria not been met at the first look, the second cohort would have been studied.

Table 13: Examples from Replicate Design Bioequivalence in Three Studies for AUC and Cmax

Study	Sample Size	AUC PE (90% CI)	Cmax PE (90% CI)
I	74	0.91 (0.84, 0.98)	0.92 (0.85, 0.99)
II	75	0.92 (0.84, 1.01)	0.94 (0.86, 1.03)
III	94	1.01 (0.91, 1.13)#	0.96 (0.86, 1.08)#
# 95% Confidence Interval			
PE: Ratio of Adjusted Geometric Means			

No gross differences in inter-subject variability were noted between formulations in these studies as demonstrated by inspection of Figure 13. Subject-by-formulation interaction variance estimates (Hauck et al., 2000) were negligible, and population and individual bioequivalence (FDA Draft Guidance, 1997, 1999a, and 1999b; FDA Guidance, 2001) were demonstrated in each study (as we will see in later Chapters for replicate cross-over designs in data sets B, R, and S).

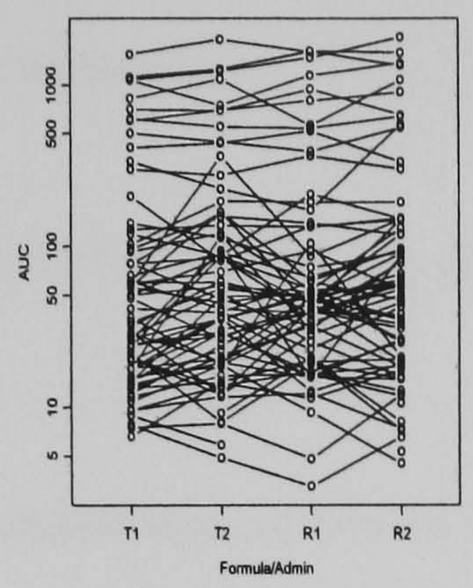
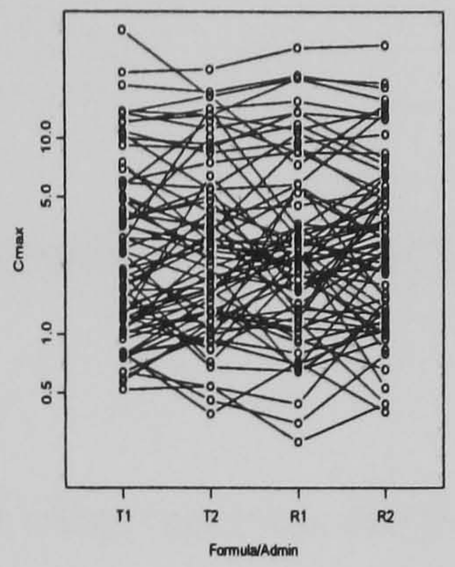
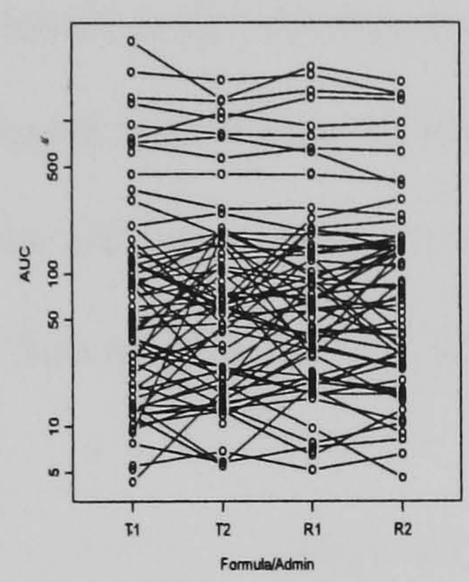
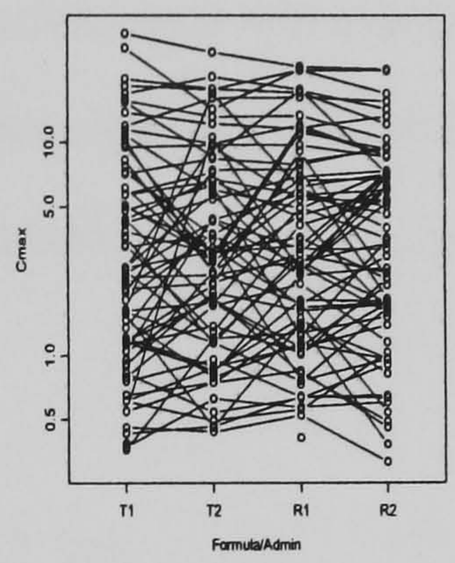
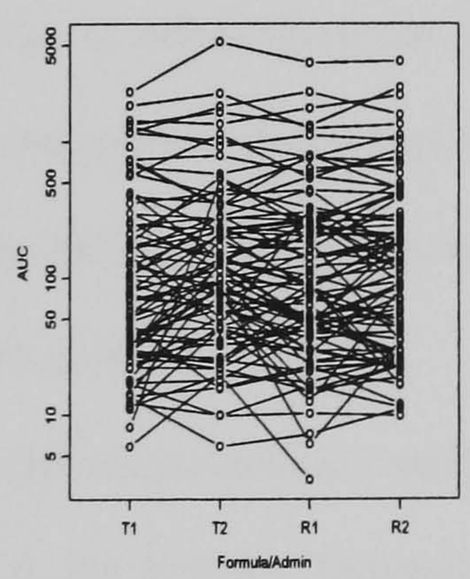
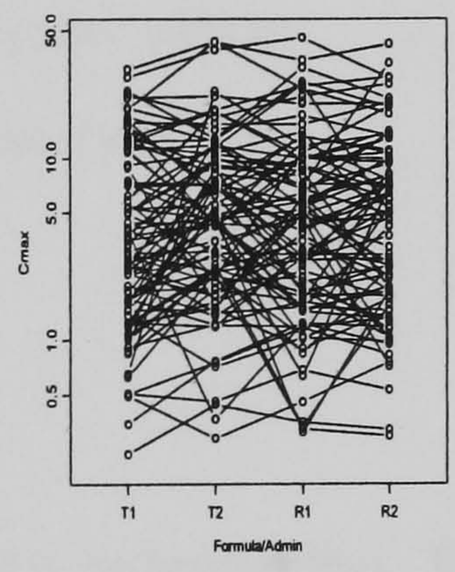
Study	AUC	Cmax
I		
II		
III		

Figure 13: AUC and Cmax for Studies I, II, and III

For the purposes of this thesis we will now concentrate upon the modelling of replicate designs in bioequivalence studies. Details of the extension of these techniques to group sequential designs will not be further explored.

2.5 Estimation Methods for Average Bioequivalence in Replicate Designs

We now turn to consideration of how to model PK data from replicate designs using method-of-moments and restricted maximum likelihood based approaches.

The following mixed model for log_e-transformed observations is sometimes used (Jones and Kenward, 1989; Vonesh and Chinchilli, 1997) in replicate designs. Let y_{tijk} be the k-th response $(k = 1, 2, \dots)$ for the j-th subject $(j = 1, 2, \dots, n_i)$ in sequence i $(i = 1, 2, \dots, s)$ in the cross-over trial administered formulation t $(t = T, R)$ and

$$y_{tijk} = \gamma_{itk} + \nu_{tj} + \mu_t + \varepsilon_{tjk} \tag{27}$$

where γ_{itk} is a vector of nuisance effects (sequence and period effects),

ν_{tj} and ε_{tjk} are independent and normally distributed with mean zero

$Var(\nu_{tj}) = \sigma^2_{Bt}$, the between-subject variance,

$Cov(\nu_{Tj}, \nu_{Rj}) = \rho\sigma_{BT}\sigma_{BR}$,

$Var(\varepsilon_{tjk}) = \sigma^2_{Wt}$, the within-subject variance,

$Cov(\varepsilon_{tjk}, \varepsilon_{tjk'}) = 0$, for $k \neq k'$. Subjects are assumed to be independent. Under this model $Var(\nu_{Tj} - \nu_{Rj}) = \sigma^2_D$, the subject-by-formulation interaction variance. Note that nuisance effects (period and sequence effects) are fitted in practice (Jones and Kenward, Chapter 4, 1989) but are omitted from the above description for the sake of clarity. The above model may be fitted using general linear models (corresponding to a method-of-moments approach), maximum likelihood or restricted maximum likelihood based procedures.

It should be noted here that one of the key provisions of the 1999 draft FDA guidance was the suggestion for what has been termed a 'public health experiment' or mandatory data collection period (Montreal, AAPS/FDA Workshop, August/September 1999). Sponsors of any bioequivalence study would be compelled to submit data from a replicate design to FDA for

approval to market. Subsequent discussion at the Advisory Committee for Pharmaceutical Science (September 1999, Washington DC) resulted in the recommendation that market access not be permitted unless average bioequivalence had been demonstrated under existing criteria unless the sponsor could convince FDA otherwise. Restricting the 'experiment' to a class of drugs such as controlled release formulation and highly variable drugs was also suggested. These proposals were adopted in the finalised guidance (FDA Guidance, 2000b), but later was changed in FDA Guidance (2002) to explicitly only allow the use of average bioequivalence for market access.

This call for data, however, required for careful and meticulous examination of existing replicate design data sets prior to beginning the 'public health experiment' so as to set realistic expectations for the exercise (Zariffa et al., 2000) and for the careful, scientific consideration of viable alternatives to the procedure developed by the FDA. The differences in statistical estimation and inference in such data sets constitutes the remainder of this Chapter.

We begin discussion of estimation procedures with method of moments and follow with discussion of maximum likelihood based procedures and the properties of estimates arising from the models. Asymptotic properties and bias in the proposed metrics will be characterised, and the content of the retrospective analysis to be presented in Section 2.7 will be defined. These findings follow from general results (Stuart et al., 1999) in the context of method of moments estimation (related to conditions on design matrices).

In complete data sets from two sequence (RTRT, TRTR), randomised, replicate designs, method-of-moments estimators may be used to calculate unbiased estimators for the parameters of interest in (27). We begin by establishing the statistical relationship between method-of-moments estimators of interest.

Theorem 2.1 *Pairwise Independence of Method-of-Moment Estimators*

In balanced, s-sequence, replicate cross-over design, with no missing data, unbiased method of moment estimators $\hat{\delta}$, M_I, M_T, *and* M_R *for* $\delta = \mu_T - \mu_R$, $\sigma_I^2 = \sigma_D^2 + \frac{\sigma_{WT}^2 + \sigma_{WR}^2}{2}$, σ_{WT}^2, *and* σ_{WR}^2 *are independent.*

Proof: Let the individual difference across formulations be denoted $I_{ij} = \bar{y}_{Tij\bullet} - \bar{y}_{Rij\bullet}$ such that

$$\hat{\delta} = \frac{1}{s}\sum_{i=1}^{s}\left[\frac{1}{n_i}\sum_{j=1}^{n_i} I_{ij}\right]$$

and

$$M_I = \frac{1}{(\sum_{i=1}^{s} n_i) - s}\sum_{i=1}^{s}\sum_{j=1}^{n_i}(I_{ij} - \bar{I}_i)^2$$

It follows that these two statistics $\hat{\delta}$ and M_I are independent based on previous results attributed to Fisher (described in Johnson et al., Vol 2, 1995 and Muirhead, 1982). Vonesh and Chinchilli (1997) show that $\hat{\delta}$ is unbiased for δ, and Chinchilli and Esinhart (1996) showed that $M_I \sim \sigma_I^2\chi_\nu^2/\nu$ where ν is the degrees of freedom associated with M_I. In a two-sequence balanced design with no missing data (RTRT, TRTR) recommended by FDA (1997-2002), $\nu = (\sum_{i=1}^{s} n_i) - s = n - 2$.

Let the individual difference within formulations for test and reference formulations be denoted

$T_{ij} = y_{Tij1} - y_{Tij2}$ and $R_{ij} = y_{Rij1} - y_{Rij2}$, respectively. Within-subject variances are estimated by

$$M_T = \frac{1}{2((\sum_{i=1}^{s} n_i) - s)}\sum_{i=1}^{s}\sum_{j=1}^{n_i}(T_{ij} - \overline{T}_i)^2$$

and

$$M_R = \frac{1}{2((\sum_{i=1}^{s} n_i) - s)}\sum_{i=1}^{s}\sum_{j=1}^{n_i}(R_{ij} - \overline{R}_i)^2$$

Chinchilli and Esinhart (1996) showed that $M_T \sim \sigma_{WT}^2\chi_\nu^2/\nu$ where ν is the degrees of freedom associated with M_T, and $M_R \sim \sigma_{WR}^2\chi_\nu^2/\nu$ where ν is the degrees of freedom associated with M_R. As M_R and M_T are derived from independent multivariate normal observations, M_R and M_T are independent under (27).

It remains to show that the estimates of within-subject variability (M_R and M_T) are pairwise independent with $\hat{\delta}$ and M_I. Here we refer the reader to the well known result based on the properties of the multivariate normal density such that if A and B are bi-variate normally distributed with non-null correlation ρ and homogeneous variance then $\frac{A+B}{2}$ and $A - B$ are independent (see Bickel and Doksum, 1977). In this context, it follows that R_{ij}, T_{ij}, and I_{ij} are mutually independent, and as subjects are independent, it follows that M_R is independent of M_I and $\hat{\delta}$, and M_T is independent of M_I and $\hat{\delta}$.□□□

We now turn to the consideration of estimation and inference in incomplete data sets. Models such as (27) are easily adaptable to more complex situations involving missing data through the use of restricted-maximum likelihood (REML) estimation (Patterson and Thompson, 1971; Harville, 1977; Laird and Ware, 1982; Brown and Kempton, 1994). Discussion of REML versus Method-of-moments estimation in this setting is an important foundation and will be thoroughly explored in this thesis.

Several previous authors have explored the modelling of such data using techniques developed for repeated measurements (Jones and Kenward, Chapter 7, 1989; Milliken and Johnson, Chapter 32, 1992; Vonesh and Chinchilli, Chapter 4, 1997; Kimanani et al., 2000a). We will concentrate on estimation in a four period, two treatment, replicate design.

Under such an approach in a four-period, two-sequence (RTRT, TRTR), replicate design study, let $\underline{y}$ be the real-valued response $4n \times 1$ vector ($p = 4$ is the number of periods). Here, there are two sequences $i = 1, 2$ corresponding to sequences RTRT and TRTR and subject within sequence $j = 1, \ldots, n_i$ where $n = n_1 + n_2$ is the overall sample size. Then,

$$\underline{y} \sim MVN(\boldsymbol{X}\underline{\beta}, \boldsymbol{\Sigma}) \tag{28}$$

where $\boldsymbol{X}$ is the known $4n \times 9$ design matrix, $\underline{\beta}$ is an 9×1 vector of fixed effect location parameters (including terms for intercept, sequence, period, and formulation, of which only 6 parameters will be estimable), and $\boldsymbol{\Sigma} = Var(\underline{y}) = \boldsymbol{Z}_1(\boldsymbol{\Omega})\boldsymbol{Z}_1' + \boldsymbol{Z}_2\boldsymbol{\Lambda}\boldsymbol{Z}_2'$ is a $4n \times 4n$ matrix of variance components. Note that $'$ denotes the transpose of a matrix, and MVN indicates a multivariate normal distribution. We now turn to the structure of $\boldsymbol{\Sigma}$ in more detail.

In matrix notation, this model can be expressed as

$$\underline{y} = \boldsymbol{X}\underline{\beta} + \boldsymbol{Z}_1\underline{u} + \boldsymbol{Z}_2\underline{e} \tag{29}$$

where $\underline{u}$ is multivariate normal with expectation $\underline{0}$ and variance-covariance matrix $\boldsymbol{\Omega}$ ($\underline{u} \sim MVN(\underline{0}, \boldsymbol{\Omega})$) and $\underline{e} \sim MVN(\underline{0}, \boldsymbol{\Lambda})$, where $\underline{u}$ is independent of $\underline{e}$. Let $\boldsymbol{\Omega}$ be defined in terms of variance-covariance components $(\sigma^2_{BR}, \omega_{RT}, \sigma^2_{BT})$ corresponding to the method-of-moments approach (Chinchilli and Esinhart, 1996) where ω_{RT} represents the covariance between test and reference observations under model (27), and let $\boldsymbol{\Lambda}$ be defined in terms of variance-covariance

components $(\sigma^2_{WR}, 0, \sigma^2_{WT})$. In addition, $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are matrices whose elements are composed of 0's or 1's and are used to assemble the covariance matrix of observations in a manner appropriate to sequence i for each subject j. Subjects are assumed to be independent.

For example, suppose a study is performed with sequences RTRT ($i = 1$) and TRTR ($i = 2$), then for two subjects with sequences RTRT and TRTR, respectively:

$\boldsymbol{\Omega}_{RTRT} = \boldsymbol{\Omega}_1$ is $\begin{pmatrix} \sigma^2_{BR} & \omega_{RT} & \sigma^2_{BR} & \omega_{RT} \\ \omega_{RT} & \sigma^2_{BT} & \omega_{RT} & \sigma^2_{BT} \\ \sigma^2_{BR} & \omega_{RT} & \sigma^2_{BR} & \omega_{RT} \\ \omega_{RT} & \sigma^2_{BT} & \omega_{RT} & \sigma^2_{BT} \end{pmatrix}$,

and $\boldsymbol{\Omega}_{TRTR} = \boldsymbol{\Omega}_2$ is $\begin{pmatrix} \sigma^2_{BT} & \omega_{RT} & \sigma^2_{BT} & \omega_{RT} \\ \omega_{RT} & \sigma^2_{BR} & \omega_{RT} & \sigma^2_{BR} \\ \sigma^2_{BT} & \omega_{RT} & \sigma^2_{BT} & \omega_{RT} \\ \omega_{RT} & \sigma^2_{BR} & \omega_{RT} & \sigma^2_{BR} \end{pmatrix}$ where $\omega_{RT} = Cov(y_{Tijk}, y_{Rijk})$.

Also, $\boldsymbol{\Lambda}_{RTRT} = \boldsymbol{\Lambda}_1 = \begin{pmatrix} \sigma^2_{WR} & 0 & 0 & 0 \\ 0 & \sigma^2_{WT} & 0 & 0 \\ 0 & 0 & \sigma^2_{WR} & 0 \\ 0 & 0 & 0 & \sigma^2_{WT} \end{pmatrix}$

and $\boldsymbol{\Lambda}_{TRTR} = \boldsymbol{\Lambda}_2 = \begin{pmatrix} \sigma^2_{WT} & 0 & 0 & 0 \\ 0 & \sigma^2_{WR} & 0 & 0 \\ 0 & 0 & \sigma^2_{WT} & 0 \\ 0 & 0 & 0 & \sigma^2_{WR} \end{pmatrix}$.

Under this approach, the *log*−likelihood function for (28) is expressed as:

$$\boldsymbol{L} = -\frac{n[\ln(2\pi)]}{2} - \frac{1}{2}\sum_{i=1}^{s} \ln \mid \boldsymbol{\Sigma} \mid -\frac{1}{2}\sum_{i=1}^{s}\sum_{j=1}^{n_i}(\underline{y} - \boldsymbol{X}\underline{\beta})'\boldsymbol{\Sigma}^{-}(\underline{y} - \boldsymbol{X}\underline{\beta}) \tag{30}$$

where $\boldsymbol{\Sigma}^{-}$ is a generalised inverse of $\boldsymbol{\Sigma}$ and $\mid \boldsymbol{\Sigma} \mid$ is the determinant of matrix $\boldsymbol{\Sigma}$ (Jones and Kenward, 1989). Equation (30) is maximised over the parameter space of $\boldsymbol{\Sigma}$ (adjusted for fixed effects $\boldsymbol{X}\underline{\beta}$ restricting the procedure to the fixed effects space, hence restricted maximum likelihood as in Patterson and Thompson, 1971) using iterative Newton-Rahpson (Lindstrom and Bates, 1988), Fisher's scoring (Jennrich and Schluchter, 1986), or EM generalised algorithms (Laird and Ware, 1982; Jennrich and Schluchter, 1986). Non-iterative procedures (known as MIVQUE0, see Milliken and Johnson, 1992) may also be used.

Restricted maximum-likelihood estimation may then be performed using a variety of software packages ($SAS^{©}$, $SPLUS^{©}$, or $GENSTAT^{©}$) to derive unbiased, best quadratic estimators for the variance components and best linear unbiased estimators for the fixed effects (Searle, 1971) only in those data sets which are balanced and have no missing data (where estimates will be the same as those derived using method-of-moments). In such data sets, the derivation of unbiased estimators for the metrics of interest follow the properties established earlier in this Chapter; however, of greater practical interest is the consideration of accurate estimation procedures in incomplete or imbalanced data sets.

Estimates from unbalanced data sets for fixed effects are referred to as empirical best linear

estimates, and estimates for random effects are empirical best linear predictors (Laird and Ware, 1982; Harville and Carriquiry, 1992). These are unbiased only in complete and balanced data sets, and they are empirical in that estimates for $\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{\Omega}}$ are derived based on least squares estimates $\underline{\hat{\beta}}$. These variance matrices, $\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{\Omega}}$, are then used to re-calculate an estimate $\underline{\hat{\beta}}$. This iterative process continues until convergence is reached (the reader is referred to Jennrich and Schluchter, 1986, for a detailed description of the procedures).

As resulting estimates for fixed and random effects do not account for the iterative estimation of $\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{\Omega}}$ and $\underline{\hat{\beta}}$, the error associated with these being empirical generalised least squares estimators is too small as this uncertainty is not taken into account, leading to confidence intervals which may be too narrow. As such would typically lead to making it easier to demonstrate average bioequivalence, this is of obvious concern in bioequivalence studies, and has not received attention in the statistical literature. As an alternative however, it should be noted that estimates from method-of-moments are best, linear, unbiased, and independent only when a balanced complete data set is available (an infrequent occurrence in bioequivalence studies, see data appearing later in this Chapter). The discrepancies between these procedures and impact on average bioequivalence inference will be explored in subsequent sections.

The resulting REML estimates in incomplete data sets, however, are not independent but are asymptotically unbiased estimators for the parameters of interest (Milliken and Johnson, Chapter 22, 1992) with known large sample variance-covariance matrix (Searle, Chapter 10, 1971). Let $\underline{\hat{\beta}}$ and $\hat{\mathbf{\Sigma}}$ be asymptotically unbiased REML estimates for $\underline{\beta}$ and $\mathbf{\Sigma}$, respectively, then the large sample variances for $\underline{\hat{\beta}}$ and $\hat{\mathbf{\Sigma}}$ are $(\boldsymbol{X}'\mathbf{\Sigma}^{-}\boldsymbol{X})^{-}$ and $-E[\frac{\partial^2 \boldsymbol{L}^{-1}}{\partial \mathbf{\Sigma} \partial \mathbf{\Sigma}'}]$, respectively with covariance $\mathbf{0}$ where $\boldsymbol{L}$ is the *log*-likelihood in expression (30).

It should be noted that, while the comparison of fixed effects in this situation is relatively well characterised (see Kenward and Roger, 1997), protection of the Type I error rate for ABE has not been established by published simulation studies (we will address this in Chapter 5), and estimates of variance are not as well characterised in small samples. In the one-way analysis of variance, it is known (Swallow and Monahan, 1984) that variance estimates are biased by a negligible amount in small samples and are unbiased in a balanced setting. However, variances estimates in the designs used for bioequivalence are not precisely characterised (Zariffa et al., 2000). The impact of this state of knowledge on the characterisation of aggregate criteria such

as PBE and IBE is of immediate concern (to be studied in Chapters 3-5), however, this does not greatly concern us for ABE unless it impacts inference. The impact of this state of knowledge on the characterisation of aggregate criteria is of immediate concern. We will remedy this situation in subsequent simulations to be described later in this thesis.

We will concentrate upon $SAS^{©}$-based mixed effect, restricted maximum likelihood (Wolfinger et al., 1994) in this chapter as this procedure is recommended in the FDA draft and final guidances (1997, 1999a, 1999b, 2001). To improve speed of convergence of these iterative procedures, $SAS^{©}$ programming uses a 'short-cut' for models like (27) based on findings from Vonesh and Chinchilli (1997). Note that based on (28) and Theorem 2.1,

$\begin{pmatrix} \bar{y}_{Rj} \\ \bar{y}_{Tj} \end{pmatrix} \sim MVN_2 \begin{pmatrix} \mu_T \\ \mu_R \end{pmatrix}, \begin{pmatrix} \sigma^2_{BR}+\frac{\sigma^2_{WR}}{2} & \omega_{RT} \\ \omega_{RT} & \sigma^2_{BT}+\frac{\sigma^2_{WT}}{2} \end{pmatrix}$, and

$\begin{pmatrix} \frac{y_{Rj1}-y_{Rj2}}{\sqrt{2}} \\ \frac{y_{Tj1}-y_{Tj2}}{\sqrt{2}} \end{pmatrix} \sim MVN_2 \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{WR} & 0 \\ 0 & \sigma^2_{WT} \end{pmatrix}$.

Hence, where $\sigma^2_{Bt} \leq \sigma^2_{Wt}$ and variance estimates are constrained (Harville, 1977) to be positive, the REML algorithms described above may fail to converge in $SAS^{©}$ (Data sets E and T are examples of this). However, by specifying starting values (method-of-moments estimates will usually serve for this purpose), using a 'PARMS' statement in $SAS^{©}$, the models can usually be made to converge. Under such circumstances, statisticians should explore several starting values to ensure that the model does not converge to a local maximum (Searle, 1971). Differences between these procedures mainly relate to the set of variance-covariance components used to estimate the matrix $\mathbf{\Omega}$. In each procedure subtle differences in the way this matrix is specified can result in differences in the resulting estimated $\mathbf{\Omega}$ matrix resulting in differing estimates for the metrics of interest in bioequivalence assessment.

The first model to be considered is referred to as 'UN' due to the 'unstructured' variance-covariance structure and 'unconstrained' parameter space used to express $\mathbf{\Omega}$ using the following set of variance-covariance parameters: $(\sigma^2_{BR}, \omega_{RT}, \sigma^2_{BT})$ corresponding to our original description. As known from Searle (1971) estimates resulting from this model should equate to a method-of-moments approach in balanced designs with no missing data. The following $SAS^{©}$ code may be used:

```
PROC MIXED METHOD=REML SCORING=50 MAXITER=200;
CLASS SEQUENCE SUBJECT PERIOD REGIMEN;
MODEL lnAUC = sequence period regimen /DDFM=SATTERTH;
```

RANDOM regimen / subject=subject type=UN;

REPEATED regimen / group=regimen type=simple;

(FDA Draft Guidance, 1997)

Recall that the structure of the resulting variance-covariance estimates for an individual subject receiving sequence RTRT, $\mathbf{\Omega}_{RTRT}$ and $\mathbf{\Lambda}_{RTRT}$, are as follows. $\mathbf{\Omega}_{RTRT}$ is $\begin{pmatrix} \sigma^2_{BR} & \omega_{RT} & \sigma^2_{BR} & \omega_{RT} \\ \omega_{RT} & \sigma^2_{BT} & \omega_{RT} & \sigma^2_{BT} \\ \sigma^2_{BR} & \omega_{RT} & \sigma^2_{BR} & \omega_{RT} \\ \omega_{RT} & \sigma^2_{BT} & \omega_{RT} & \sigma^2_{BT} \end{pmatrix}$ where

$$\omega_{RT} = \omega_{TR} = Cov(y_{Tijk}, y_{Rijk})$$

and $\mathbf{\Lambda}_{RTRT}$ is $\begin{pmatrix} \sigma^2_{WR} & 0 & 0 & 0 \\ 0 & \sigma^2_{WT} & 0 & 0 \\ 0 & 0 & \sigma^2_{WR} & 0 \\ 0 & 0 & 0 & \sigma^2_{WT} \end{pmatrix}$. Note that $\mathbf{\Omega}$ may not be non-negative definite.

The second model is referred to as 'CSH' due to the variance-covariance structure and parameter space used to express $\mathbf{\Omega}$ using the following set of variance-covariance parameters $(\sigma^2_{BR}, \rho, \sigma^2_{BT})$ and using the following $SAS©$ code:

PROC MIXED METHOD=REML SCORING=50 MAXITER=200;

CLASS SEQUENCE SUBJECT PERIOD REGIMEN;

MODEL $lnAUC$ = sequence period regimen /DDFM=SATTERTH;

RANDOM regimen / subject=subject type=CSH;

REPEATED / group=regimen type=simple;

(FDA Draft Guidance, 1997)

The procedure 'CSH' is sometimes referred to as 'Constrained REML' as $\mathbf{\Omega}$ is constrained to be non-negative definite (Harville, 1977) due to the constraint placed upon ρ, and the $log-$likelihood is maximised over the constrained parameter space (though this procedure may be modified to be unconstrained by adding the statement 'PARMS /NOBOUND;' to the $SAS©$ PROC MIXED procedure above).

The third model is referred to as 'FA0(2)' due to the variance-covariance structure and parameter space used. $\mathbf{\Omega}$ is re-expressed as based upon the set of variance-covariance parameters: $(\sigma_{BR}, \sigma_D, \sigma_{BT})$ with adjustment to $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ as appropriate. This procedure could also be referred to as 'Constrained REML' as $\mathbf{\Omega}$ is constrained to be non-negative definite using $SAS©$ code as follows:

PROC MIXED METHOD=REML SCORING=50 MAXITER=200;

CLASS SEQUENCE SUBJECT PERIOD REGIMEN;

MODEL $lnAUC$ = sequence period regimen /DDFM=SATTERTH;

RANDOM regimen / subject=subject type=FA0(2);

REPEATED / group=regimen type=simple;

(FDA Draft Guidance, 1999b; FDA Guidance, 2001)

Theoretically the three approaches (UN, CSH, and FA0(2)) should result in the same matrix $\mathbf{\Omega}$ as the functional forms are mathematically equivalent (as can be seen from the above). However, the matrices $\mathbf{\Omega}$ can differ based on the parameterization of the model space corresponding to whether the covariance is denoted directly (UN: ω_{RT}), denoted as a function of the correlation and variances (CSH: $\rho\sigma_{BT}\sigma_{BR}$), or denoted as a function of the difference in between-subject variances (FA0(2): $\frac{\sigma^2_{BR}+\sigma^2_{BT}-\sigma^2_D}{2}$). The constrained REML estimators for σ^2_D using CSH have been shown (Zariffa et al., 1998; Endrenyi and Tothfalusi, 1999; Kimanani et al., 2000a) to be positively biased when estimates are close to zero, and this should effect the assessment of average and individual bioequivalence. However, evidence to date of positive bias has been empirical or simulation based (Endrenyi and Tothfalusi, 1999) and has not been observed to effect inference (Hauck et al., 2000). We will study the bias and impact on inference further using retrospective analysis and simulation.

We propose an alternative model, directly estimating the parameters of interest, based on the use of a random-intercept, random-slope model based on the cross-over model of Jones and Kenward (1989) and based on techniques described in Milliken and Johnson (Chapter 22, 1992). Let y_{ijkl} be the l-th ($l = 1, 2$) replicated $\log_e$-transformed j-th period's observation ($j = 1, 2, 3, 4$) for the k-th subject ($k = 1, 2, \ldots, n_i$) in the i-th sequence group ($i = 1, 2$). Then

$$y_{ijkl} = \lambda_i + (\mu + \nu_{k(i)}) + \pi_j + (\beta + \xi_{k(i)})\tau_{d[i,j]} + \varepsilon_{d[ijkl]} \tag{31}$$

where μ is the grand mean,

λ_i, π_j, and $\tau_{d[i,j]}$ are fixed effects for sequence, period, and formulation, respectively,

$\nu_{k(i)}$, $\xi_{k(i)}$ and $\varepsilon_{d[ijkl]}$ are random effects which are normally distributed and independent with mean zero, $Var(\nu_{k(i)})=\sigma^2_B$, the pooled (across formulations) between-subject variance, $Var(\xi_{k(i)})=\frac{\sigma^2_D}{2}$, half the subject-by-formulation interaction, and

$Var(\varepsilon_{d[ijkl]})=\sigma^2_{Wt}$, the within-subject variance for test and reference formulations. In matrix

notation, the model is again re-expressed as

$$\underline{y} \sim MVN(\boldsymbol{X}\underline{\beta}, \boldsymbol{\Sigma}) \tag{32}$$

where $\boldsymbol{X}$ is the design matrix as above, $\underline{\beta}$ is an 9×1 vector of fixed effect location parameters as above, and $Var(\underline{y}) = \boldsymbol{Z\Omega Z'} + \boldsymbol{\Lambda}$ is a $4n \times 4n$ matrix of variance components. $\boldsymbol{\Omega}$ is composed in terms of the set $(\sigma_B^2, \sigma_D^2/2)$, and $\boldsymbol{\Lambda}$ is defined as above. In $SAS^{©}$, this can be analysed using the following code:

```
PROC MIXED METHOD=REML SCORING=50 MAXITER=200;
CLASS SEQUENCE SUBJECT PERIOD REGIMEN;
MODEL lnAUC = sequence period regimen /DDFM=SATTERTH;
RANDOM intercept regimen / subject=subject type=simple;
REPEATED / group=regimen type=simple;
```

Here, for a subject receiving sequence RTRT, $\boldsymbol{\Omega}_{RTRT}$ is $\begin{pmatrix} \sigma_B^2+\frac{\sigma_D^2}{2} & \sigma_B^2 & \sigma_B^2+\frac{\sigma_D^2}{2} & \sigma_B^2 \\ \sigma_B^2 & \sigma_B^2+\frac{\sigma_D^2}{2} & \sigma_B^2 & \sigma_B^2+\frac{\sigma_D^2}{2} \\ \sigma_B^2+\frac{\sigma_D^2}{2} & \sigma_B^2 & \sigma_B^2+\frac{\sigma_D^2}{2} & \sigma_B^2 \\ \sigma_B^2 & \sigma_B^2+\frac{\sigma_D^2}{2} & \sigma_B^2 & \sigma_B^2+\frac{\sigma_D^2}{2} \end{pmatrix}$

Classical model building-maximum likelihood based testing procedures (Neyman-Pearson type testing procedures described in Milliken and Johnson, Chapter 1, 1992) to aid in this assessment however are not well established in data sets as small as those usually encountered in bioequivalence studies and should be applied with caution. However, techniques for model discrimination are available in the context of mixed modelling (see Lindsey and Jones, 1997 and Lindsey et al., 1999). The Schwarz Bayesian criterion (SBC; Schwarz, 1978) and the Akaike Information criteria (AIC; Akaike, 1973) may be applied to discriminate between models which differ in variance-covariance structure, and we will study what these criteria determine concerning the models later in this chapter.

2.6 Properties of the Estimated Metrics and Inferential Procedures for ABE Assessment

Comparisons between the estimated means $\hat{\mu}_T$-$\hat{\mu}_R$ are (Jones and Kenward, 1989) normally-distributed with mean μ_T-μ_R and variance of

$((\sigma_B^2 + \sigma_W^2) + (\sigma_B^2 + \sigma_W^2) - 2(\rho)(\sqrt{\sigma_B^2 + \sigma_W^2})(\sqrt{\sigma_B^2 + \sigma_W^2}))/n = 2(\sigma_W^2)/n$ in balanced two-

period cross-over designs with no missing data and n subjects. Estimates of variance may be derived using method-of-moments or REML estimation, and these estimates are unbiased (or asymptotically unbiased in the REML case) for the true variances. Tests of fixed effects are exact under the Huyhn-Feldt condition, and may be constructed in the usual fashion (see Chapter 1) for the assessment of average bioequivalence (FDA Guidance 1992, 2001) in balanced, complete two-period cross-over data sets using method-of-moments.

In bioequivalence studies, the Huyhn-Feldt condition for variances across formulations may not be applicable, and in a replicate design, the different variance estimates of each formulation are estimable. Vonesh and Chinchilli (1997) present methods for the estimation of total, between, and within-subject variances in replicate and cross-over designs. Here we note only that careful consideration of the covariance between observation must be implemented when analysing such data so as to prevent the classic errors associated with the analysis of cross-over data (Senn, 1993; Senn, 2002). Note that estimates σ_T^2 and σ_R^2 in a two-period cross-over and estimates $\sigma_t^2 = \sigma_{Bt}^2 + \frac{\sigma_{Wt}^2}{2}$ (for $t = T, R$ representing Test and Reference formulations) in replicate designs should be correlated in a cross-over design (Jones and Kenward, 1989).

Under the approaches to analysis of the replicate design described previously in this Chapter, it can be shown (Vonesh and Chinchilli, Chapter 4, 1997) that the variance for $\hat{\mu_T} - \hat{\mu_R}$ is equal to $(\sigma_{BT}^2 + \sigma_{BR}^2 - 2\rho\sigma_{BT}\sigma_{BR} + ((\sigma_{WT}^2 + \sigma_{WR}^2)/2))/n = (\sigma_D^2 + ((\sigma_{WT}^2 + \sigma_{WR}^2)/2))/n$ in a balanced design with n subjects. In a two-period cross-over, the same model is used under the assumptions that inter-subject variance is homogeneous between formulations, that ρ is equal to unity, and that intra-subject variance is homogeneous between formulations (FDA Guidance, 1992). It can be shown (Vonesh and Chinchilli, Chapter 4, 1997) that the variance for $\hat{\mu_T} - \hat{\mu_R}$ in a two-period cross-over is equal to $(\sigma_{BT}^2 + \sigma_{BR}^2 - 2\rho\sigma_{BT}\sigma_{BR} + \sigma_{WT}^2 + \sigma_{WR}^2)/n = (\sigma_D^2 + \sigma_{WT}^2 + \sigma_{WR}^2)/n$ in a balanced design with n subjects.

Small sample inference for average bioequivalence (FDA Guidance, 1992-2001; see also Chapter 1 and Section 2.1) where the comparison of interest involves $\mu_T - \mu_R$ in a balanced or unbalanced replicate design involves explicitly only fixed effects (Jones and Kenward, 1989; Kenward and Roger, 1997). In this context, it is easy to show (Vonesh and Chinchilli, 1997) that the

variance of $\hat{\mu}_T - \hat{\mu}_R$ is normally distributed with mean $\mu_T - \mu_R$ and variance of

$$\frac{\sigma_D^2 + \frac{\sigma_{WT}^2 + \sigma_{WR}^2}{2}}{n}$$

in a complete data set. Inference for $\mu_T - \mu_R$ involves the estimated variances based on the equation

$$\frac{\hat{\sigma}_D^2 + \frac{\hat{\sigma}_{WT}^2 + \hat{\sigma}_{WR}^2}{2}}{n}$$

which is assumed (in the case of missing data) to be centrally-χ^2 distributed. Satterthwaite's (1941) or Kenward and Roger's approximations (1997) to the degrees of freedom may be applied (these degrees of freedom are equivalent for one degree of freedom contrasts). Simulations for other study designs have shown (Kenward and Roger, 1997) that regardless of imbalance in the data set, assessment of mean differences appears to maintain a nominal Type I error rate. However, simulations referenced in Kenward and Rogers' (1997) previous work did not consider a replicate design, and work in this thesis will extend their findings to characterise the properties of the estimates arising from such models using simulation (Chapter 5). We first will use a data base of 51 replicate design data sets to investigate the properties of the models involved.

In balanced, two-sequence (RTRT, TRTR), replicate cross-over design, with no missing data, unbiased method of moment estimators $\hat{\delta}$, M_I, M_T, and M_R for $\delta = \mu_T - \mu_R$, $\sigma_I^2 = \sigma_D^2 + \frac{\sigma_{WT}^2 + \sigma_{WR}^2}{2}$, σ_{WT}^2, and σ_{WR}^2 are known to be independent as shown in Theorem 2.1, where the individual difference across formulations be denoted $I_{ij} = \bar{y}_{Tij\bullet} - \bar{y}_{Rij\bullet}$ such that

$$\hat{\delta} = \frac{1}{s}\sum_{i=1}^{s}\left[\frac{1}{n_i}\sum_{j=1}^{n_i} I_{ij}\right]$$

and

$$M_I = \frac{1}{(\sum_{i=1}^{s} n_i) - s}\sum_{i=1}^{s}\sum_{j=1}^{n_i}(I_{ij} - \bar{I}_i)^2$$

Vonesh and Chinchilli (1997) show that $\hat{\delta}$ is unbiased for δ, and Chinchilli and Esinhart (1996) showed that $M_I \sim \sigma_I^2\chi_\nu^2/\nu$ where ν is the degrees of freedom associated with M_I. In a two-sequence (RTRT, TRTR) balanced replicate design with no missing data, $\nu = (\sum_{i=1}^{s} n_i) - s =$

$n-2$, and it is easy to show that

$$\hat{\delta} \mp t_{n-2}(0.95)\sqrt{\frac{M_I}{n}}$$

is a 90% confidence interval for $\delta = \mu_T - \mu_R$ (Vonesh and Chinchilli, 1997). This noted, we only rarely find balanced replicate designs with no missing data in such studies (see Section 2.7.1). Little focussed attention has been placed on the statistics of this situation to date.

It is known (Searle, 1971) that REML estimates for the moments of interest $\delta, \sigma^2_D, \sigma^2_{WT}, \sigma^2_{WR}$ are asymptotically normally distributed with known variances. The large sample variance for $\underline{\hat{\beta}}$ and $\hat{\boldsymbol{\Sigma}}$ are $(\boldsymbol{X}'\boldsymbol{\Sigma}^{-}\boldsymbol{X})^{-}$ and $-E[\frac{\partial^2 \boldsymbol{L}^{-1}}{\partial \boldsymbol{\Sigma} \partial \boldsymbol{\Sigma}'}]$, respectively with covariance $\mathbf{0}$ where $\boldsymbol{L}$ is the *log-*likelihood in expression (30).

Theorem 2.2 *An Asymptotic Confidence Interval for* δ

$\hat{\delta}$ will be normally distributed in the limit with expected value δ and large sample variance $\underline{l}'(\boldsymbol{X}'\boldsymbol{\Sigma}^{-}\boldsymbol{X})^{-}\underline{l}$ where $\underline{l}'$ is a vector such that $\underline{l}'\underline{\hat{\beta}}$ is asymptotically unbiased for $\underline{l}'\underline{\beta} = \delta$.

Proof: Let $g(\underline{\hat{\beta}}) = \underline{l}'\underline{\hat{\beta}}$ be a linear function of $\underline{\hat{\beta}}$ such that $g(\underline{\beta}) = \delta$. Then under Theorem 3.3.A of Serfling (1980), $\left.\frac{\partial g}{\partial \hat{\mu_T}}\right|_{g=\mu_T} = 1$ and $\left.\frac{\partial g}{\partial \hat{\mu_R}}\right|_{g=\mu_R} = -1$.

Then by application of Theorem 3.3.A (Serfling, 1980), it is found that $g(\hat{\mu_T}, \hat{\mu_R})$ is asymptotically normally distributed with expected value $g(\mu_T, \mu_R)$ and variance $\underline{D}\boldsymbol{\Sigma}_l\underline{D}'$ where $\underline{D} = (0, 0, , , , , , 1, -1)$ and where $\boldsymbol{\Sigma}_l = \boldsymbol{X}'\boldsymbol{\Sigma}^{-}\boldsymbol{X})^{-}$. The proof then follows by matrix multiplication.□□□

Thus, substituting the estimated large sample variances $\hat{\boldsymbol{\Sigma}}$ into this equation, it is easy to see that

$$\underline{l}'\underline{\hat{\beta}} \mp Z(0.95)\sqrt{\underline{l}'(\boldsymbol{X}'\boldsymbol{\Sigma}^{-}\boldsymbol{X})^{-}\underline{l}}$$

is an asymptotic 90% confidence interval for $\delta = \mu_T - \mu_R$ (where $Z(0.95)$ is the 95th quantile of the normal distribution).

In many situations for average bioequivalence, however, sample size will be lower than $20-30$ calling into question the validity of an asymptotic procedure. Kenward and Roger's (1997) and Satterthwaite's (1941) techniques (these result in equivalent estimates in 1 degree of freedom contrast such as those of concern in ABE testing) may be used in this situation to develop a small sample confidence interval using an approximated t-distribution. Here, it is proposed that

a confidence interval as follows be derived for ABE assessment using the approach Giesbrecht and Burns (1985). Let

$$\underline{l}'\hat{\underline{\beta}} \mp t_{\hat{\nu}}(0.95)\sqrt{\underline{l}'(\boldsymbol{X}'\boldsymbol{\Sigma}^{-}\boldsymbol{X})^{-}\underline{l}}$$

where $t_{\hat{\nu}}(0.95)$ is the 95th quantile of a t-distribution with

$$\hat{\nu} = \frac{(\sum_i (c_i^2\hat{\sigma}_i^2)/n_i)^2}{\sum_i (c_i^4\hat{\sigma}_i^4/n_i^2(n_i-1))}$$

with c_i representing the coefficient corresponding to $\underline{l}'\boldsymbol{X}'$ and n_i representing the number of observations contributing to $\hat{\sigma}_i^2$.

As we will see (Chapter 5), the constrained REML procedure recommended by FDA Guidance (2001) using Satterthwaite (1941) degrees of freedom for ABE testing in replicate designs results in biased estimates for variance components on occasion; however, it uniformly constrains the rate of Type I error (of more immediate concern to Regulators and Consumers) to be less than 5% in ABE testing. Thus the FDA Guidance (2001) tacitly acknowledges that 'While all models are wrong, some are useful.' If Kenward and Roger's (1997) approach to estimation of the degrees of freedom is used, as currently implemented in $SAS^{©}$, the degrees of freedom are the same as those found using Satterthwaite's procedure, however, the variance estimate for the confidence interval is inflated using the approach of Harville and Jeske (1992) to account for uncertainty introduced by iterative estimation in the fixed and random effects from the mixed modelling procedure to provide a confidence interval with at least 90% coverage probability. This results in confidence bounds which are slightly larger than those found when such an approach is not used and leads to slightly more conservative Type I error rates than those observed using the Satterthwaite option in $SAS^{©}$. As these rates already protect public health risk (as established later in this thesis using simulation), we conclude that Kenward and Roger's (1997) procedure may also be applied to test for average bioequivalence and is protective of the Type I error rate for ABE. For the purposes of the retrospective analysis presented in this Chapter, Satterthwaite's degrees of freedom (and option is $SAS^{©}$) was utilised as this is the recommended approach at present (cf. FDA Guidance, 2001) and is less conservative than the Kenward-Roger (1997) approximation.

We will now turn to the exploration of data using these techniques. Simulations to answer

questions raised by these analyses will be reported in a Chapter 5 of this thesis.

2.7 Retrospective Analysis

We now have straightforward means of assessing the average bioequivalence in replicate designs between formulations using a variety of techniques. Several issues will be assessed by retrospective analysis in Section 2.7:

1. Model Discrimination among REML alternatives

2. Differences in REML model estimates

3. Use of REML versus Method-of-Moments estimation for ABE assessment under recommendations of relevant regulatory guidance and using the constrained and unconstrained REML asymptotic procedures described above.

Unfortunately, as we will see subsequently in this Chapter, complete data sets are a rarity in bioequivalence studies, and method-of-moments estimation is of limited utility in such situations.

2.7.1 Data

Tables summarising the data to be analysed in retrospective analysis may be found in Tables 32-33. Data analysed in this retrospective analysis are from replicate design studies performed at GlaxoSmithKline Pharmaceuticals and other studies posted to the FDA website (http://www.fda.gov/cder/guidance/). All studies performed by GlaxoSmithKline were conducted in compliance with good clinical practice and were conducted according to a study protocol approved unconditionally prior to study start by an independent ethics review board. AUC and Cmax data were derived in each study period based on non-compartmental pharmacokinetic methods and were analysed using *SAS*© version 6.12.

Of the 51 data sets (Table 32: A through ZF) from replicate designs in previous studies, 31 data sets contained no missing data. Of the 20 data sets remaining, 14 data sets contained at least one missing AUC and Cmax observation. Five data sets were missing at least one observation for AUC but none for Cmax, and one data set (likely a data entry error) was missing an Cmax observation but had all observations for AUC. Data sets D, F, G, I1, I2, J, K1-3, M, N1, W1-6, and X are balanced (i.e. contain no missing data and equal numbers of

subjects in each sequence).

Twenty-eight data sets were randomised to 4 sequences (26 of which were TTRR, RRTT, RTTR, TRRT and the other 2 data sets were TTRR, RRTT, RTRT, TRTR), and 21 data sets were randomised to 2 sequences (19 of which were RTTR, TRRT and only two of which were RTRT, TRTR). Two data sets had 5 sequences of treatment administration.

Only twenty-eight data sets were available with corresponding demographic information (Table 33: ethnicity, age, weight, height.) Of these approximately one-third (9 of 28 data sets) were composed of an all male population. Fourteen of the 28 data sets were Caucasian, with the remaining data sets composed of at least one black or other (Oriental, Indian) subject. Height, weight, and age were standardised according to each study (where each subject served as their own control). Nineteen of 28 data sets with information available on gender had at least one female subject; however, only data sets C1, C2, F, Q1, Q2, and R had at least as many females as males.

2.7.2 REML Model Discrimination

We begin with discussion of REML model discrimination in the data base of replicate designs. REML model discrimination Akaike (AIC) and Schwarz Bayesian (SBC) criteria are listed in Tables 34 (AUC) and 36 (Cmax) for the four models (UN=Unstructured, CSH=Hetersceastic compound symmetry, FA0(2)=First-order analytic, RIS=Random-intercept and slope) discussed in Section 2.2. The corresponding value of the $log-$likelihood function given the estimates of the final converged model (discussed earlier in this Chapter) for AUC and Cmax are listed in Tables 35 and 37, respectively. Increasing AIC or SBC are indicative of better model fit.

Inspection reveals that, while the AIC are larger for the RIS model relative to the others for the majority of data sets (suggestive of superior model fit), the change in criteria was usually less than or equal to 1 (the change in model degrees of freedom) for AIC. This also held true if other models were compared to the RIS model. Thus this criterion implies that the models are indistinguishable in terms of their performance relative to the value of the $log-$likelihood.

For the SBC, however, the RIS model appeared to fit the data slightly better than the other models. For AUC, 33 data sets appeared to have improved fit in comparison to the FA0(2) and CSH models and for 29 data sets relative to the unstructured model as indicated by an increase

in the criterion of greater than or equal to $\frac{1}{2log(n)}$ (the change in model degrees of freedom accounting for sample size). For Cmax, the RIS model SBC exceeded the SBC of the FA0(2), CSH, and UN models by a difference greater than or equal to $\frac{1}{2log(n)}$ in 37, 36, and 31 data sets respectively. However, it should be noted that the increases were not that much greater than $\frac{1}{2log(n)}$ indicating that any improvement in RIS model fitting were very slight.

Exploring this further, we note that the AIC difference for nine AUC data sets and six Cmax data sets were greater than 1 for the UN model relative to the RIS model (indicating a slight improvement in fit for the UN covariance structure). However, in four of these AUC data sets (F, K3, ZD1, ZD2) and in three of the Cmax data sets (ZC2, ZD1, ZD2) the $\hat{\sigma}^2_D \leq 0$ implying that the $\hat{\Omega}_i$ were not non-negative definite. For the SBC, two AUC and Cmax data sets (ZD1, ZD2) had a difference in UN SBC relative to RIS SBC of greater than $\frac{1}{2log(n)}$ indicative of improved fit, however, in both data sets $\hat{\sigma}^2_D \leq 0$.

It should be noted that the values of AIC, SBC, and $log-$likelihood were identical in all data sets for the FA0(2) and CSH models, with the exception of data set E (to be examined in detail later in this Chapter). Fitted AIC and SBC for the UN model did not appear to result in distinguishable models relative to the CSH and FA0(2) models for the majority of AUC data sets (43 and 46 data sets, respectively) and Cmax data sets (48 and 48 data sets, respectively). Of the seven data sets for AIC and the four data sets for SBC where a difference in AIC of greater than 1 (AIC) or $\frac{1}{2log(n)}$ for SBC was noted for the UN model, all cases (F, K3, L2, U, X, ZD1, ZD2 for AIC and F, K3, ZD1, ZD2 for SBC) had $\hat{\sigma}^2_D \leq 0$ implying that the $\boldsymbol{\Omega}_i$ were not non-negative definite. Similar results were evident for Cmax in data sets ZD1 and ZD2.

Thus, it was found that values of the $log-$likelihood function, and information-statistics based upon them, did not appear strikingly different between models. Where the UN model performs better, in a qualitative sense, the estimates of σ^2_D were frequently negative, greatly complicating interpretation (to be discussed further in Section 2.8). These findings in combination suggest that the REML models' performance is not distinguishable in practice but that care should be exercised in choice of best model based on review of the arising estimates.

2.7.3 Point Estimates from Method-of-Moments and REML Estimation

We now turn to consideration of the estimates arising from these REML models and from method-of-moments estimation.

Method-of-moments and REML based estimates may be found in Tables 38-47. Note that 18 data sets for AUC (Data sets A, B, C1, F, K1, K3, L1, L2, N1, S, U, W1, W2, W3, X, ZD1, ZD2, ZD3) and 25 data sets for Cmax (Data Sets A, B, C1, E, J, K1, K2, K3, L1, L2, R, S, V, W1, W2, W6, Y, ZC1, ZC2, ZC3, ZD1, ZD2, ZD3) had $\hat{\sigma}_D^2 < 0$ when estimated using method-of-moments. REML (UN) estimates were similarly negative, save for data set S (AUC) and ZA (Cmax). In 13 data sets (A, B, C, K1, K3, L1, L2, S, W1, W2, ZD1, ZD2, ZD3) $\hat{\sigma}_D^2 < 0$ for both AUC and Cmax method-of-moments analyses.

Differences in estimates were minor in number for method-of-moments estimates relative to REML estimates from the UN model for the difference in formulation means ($e^{\hat{\mu}_T - \hat{\mu}_R}$) for complete and balanced data sets. However, within-subject and between-subject variance estimates were not homogeneous. Differences however were slight and likely due to numerical error in method-of-moment and REML procedures. It will be of interest subsequently to see whether the differences in procedures result in differences in inferences with any frequency.

Model estimates for $e^{\hat{\mu}_T - \hat{\mu}_R}$ appeared homogeneous for AUC and Cmax between the FA0(2) and CSH models (see Tables 42-45), except for data set E (Cmax) which fails to converge in REML unless extraordinary steps are taken (in the $SAS^{©}$ code, starting values must be pre-specified for the variance components). Estimates for $\hat{\sigma}_{WT}^2$ were also homogeneous, except for data set R (AUC) and E and W4 (Cmax), as were estimates for $\hat{\sigma}_{WR}^2$ (with the data sets E and W4 displaying differences between FA0(2) and CSH for AUC and Cmax).

Between-subject variances were more heterogeneous. Estimates for $\hat{\sigma}_{BT}^2$ differed for data sets C2, D, and R in AUC data, however, $\hat{\sigma}_{BR}^2$ in the FA0(2) analysis were lower than $\hat{\sigma}_{BR}^2$ in the CSH analysis in data sets C2, D, G, H, I2, M, N2, O1, O2, Q1, Q2, R, T, V, W4, W5, W6, Y, ZA, ZB, ZC1, ZC3, ZD4, ZE1, ZE2, ZE3 (26 of the data sets for AUC). For Cmax data, estimates for $\hat{\sigma}_{BT}^2$ differed for data sets E, F, and W4, however, $\hat{\sigma}_{BR}^2$ in the FA0(2) analysis were lower than $\hat{\sigma}_{BR}^2$ in the CSH analysis in data sets C2, D, G, H, I1, I2, M, N1, N2, O1, O2, P, Q1, Q2, T, W3, X, ZA, ZB, ZD4, ZE1, ZE2, ZE3, and ZF (24 of the data sets for Cmax). In 3 Cmax data sets (E, F, S), estimates for $\hat{\sigma}_{BR}^2$ in the FA0(2) analysis were higher than $\hat{\sigma}_{BR}^2$ in the CSH

analysis. In the majority of cases, this contributed to estimates for $\hat{\sigma}_D^2$ for CSH being greater than $\hat{\sigma}_D^2$ estimates in FA0(2) analyses, see Figure 14. It will be of interest to see whether this decrease in estimated variation impacts inference.

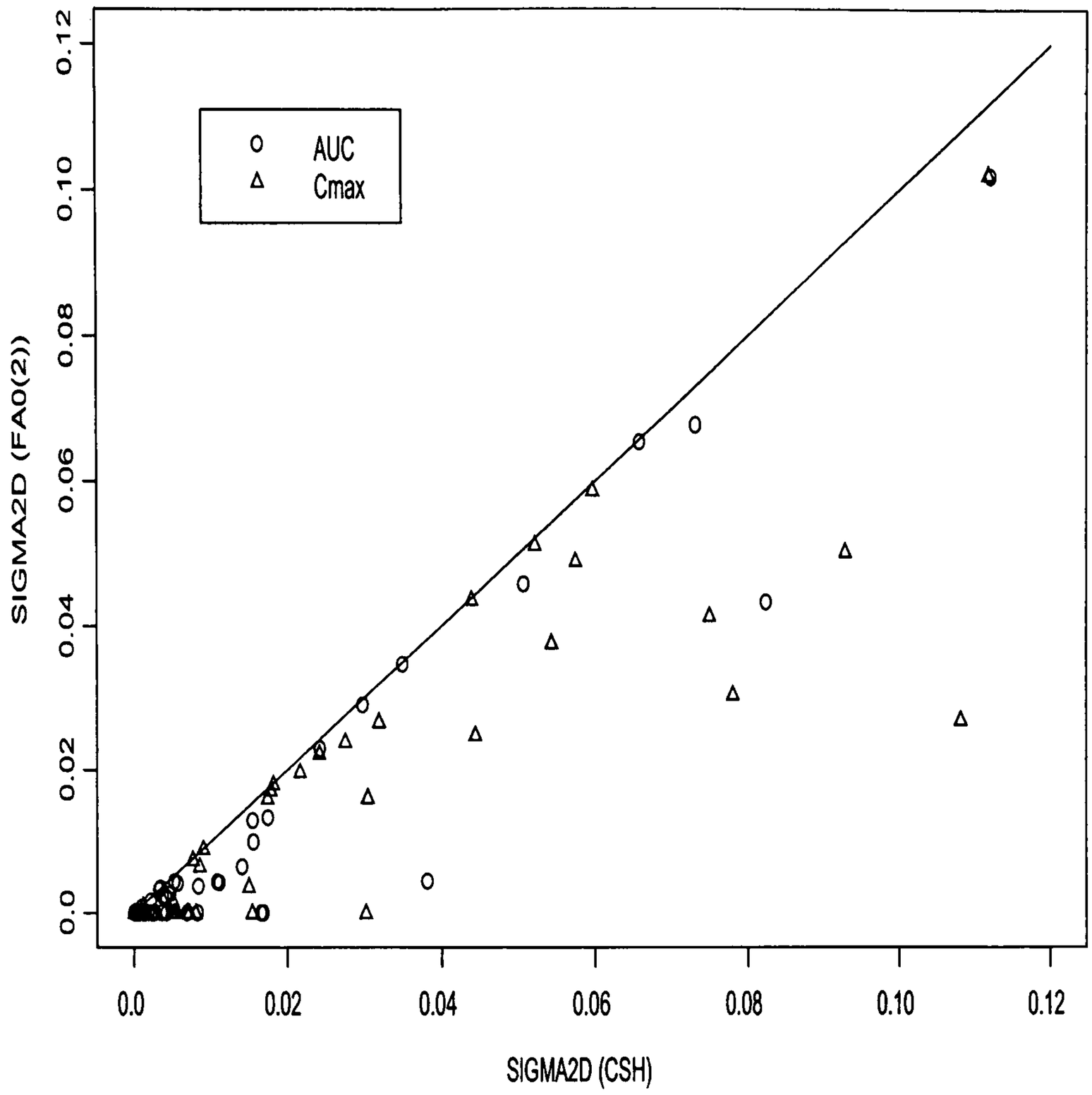


Figure 14: Subject-by-Formulation Interaction Variation for AUC and Cmax

Estimates for the RIS analysis appeared similar to the other procedures. No clear pattern was evident, and the magnitude of variance estimates were similar to other procedures (except in that the estimate for $\hat{\sigma}_D^2$ is constrained to be greater than or equal to zero in accordance with the other constrained REML procedures).

Therefore, while the REML models appear to perform equally well in practical application. Estimates from some models may not be 'meaningful' - in the classical sense of true variance estimates being positive. We will discuss the interpretation of the null and negative variance estimates in Section 2.8 and assess whether the difference in estimates between models results in differential inference in Section 2.7.4.

2.7.4 Average Bioequivalence Assessment from Method-of-Moments and REML Estimation

Inference reached based upon the REML models agreed closely with that of the method-of-moments procedure (see Tables 48 through 53). Point estimates agree closely between procedures, and in general, although variance estimates may differ to some small extent between models (see previous Section for details), in only very few cases did inference differ between procedures.

In particular, estimates for the ratio of geometric means, lower and upper 90% bounds appeared nearly identical between the CSH and FA0(2) procedures (as would be expected based on their model assessment results and the properties of the constrained REML estimation procedure) though careful inspection of the REML results revealed a very subtle decrease in the width of the confidence interval due to a decrease in the $\hat{\sigma}_D^2$ for the FA0(2) model. Method-of-moments and Unstructured REML model upper and lower bounds appeared narrower than the corresponding CSH/FA0(2) bounds as would be expected in 'unconstrained' procedures when $\hat{\sigma}_D^2$ can be less than or equal to 0. Interestingly, no clear pattern was evident in the random-intercept and slope models limits, though in most cases the limits were similar to the CSH and FA0(2) limits (as would be expected in that the RIS model is also a constrained REML procedure).

In data set L2 for AUC, the ratio of geometric means (see Table 48) was 0.871 for the method-of-moments procedure versus approximately 0.864-0.865 for the REML modelling procedures

(not to be unexpected in such an imbalanced data set.) Note however, that while the lower 90% bound was 0.808 and 0.801 for the method-of-moments and Unstructured REML models, respectively, the lower 90% bound was 0.792 for the CSH, FA0(2), and RIS models. Thus while one set of models and procedures indicates that average bioequivalence had been demonstrated, the REML models constraining $\hat{\sigma}_D^2 \geq 0$ disagreed. Indeed, results of the method-of-moments and unstructured REML modelling procedures indicated that $\hat{\sigma}_D^2 < 0$ while the other procedures constrained it to be null. Final conclusions in such a data set are problematic as the true value is not known, and REML model discrimination statistics (see earlier discussion in this Chapter) were not informative.

In data set ZA for AUC, the upper 90% bound for the method-of-moment procedure fell just above 1.25 while the REML models indicated an upper bound of just below 1.25. Estimates for the ratio of geometric means and $\hat{\sigma}_D^2$ were slightly higher for the method-of-moments model relative to the REML counterparts. Such results are not unexpected in such small unbalanced data sets.

Of the remainder where procedures agree, most of the data sets demonstrated average bioequivalence for AUC regardless of procedure. For AUC, 43 data sets demonstrated ABE while six data sets did not. Of these six, five data sets failed ABE due to a lower bound below the 0.80 cut-off (Data Sets G, I1, I2, T, ZD1; see Table 49), and one data set failed due to an upper bound in excess of the 1.25 cut-off (Table 50: Data Set Q2).

In data set Y for Cmax, the method-of-moments, CSH, and FA0(2) procedures indicated that average bioequivalence was not demonstrated while the unstructured and random-intercept and slope models indicated that it was. While the within-subject variance estimates for the reference formulation appeared slightly lower for the Unstructured and RIS models relative to the other procedures, no clear cause of these discrepancies were observed in review of the data.

Of the remaining Cmax data sets, most ($n = 36$; see Tables 51 through 53) demonstrated average bioequivalence. Nine data sets failed to demonstrate ABE due to a lower bound falling below the 0.80 cut-off (Data Sets A, G, I1, I2, L2, P, T, ZD1, ZE1; see Table 52) while six data sets failed to demonstrate ABE due to an upper bound in excess of the 1.25 cutoff (Data Sets F, O2, Q2, ZA, ZB, ZF; see Table 53).

2.8 Discussion and Findings

For highly variable drug products (intra-subject coefficient of variation greater than 30%), replicate design studies are clearly the study design of choice in demonstrating average bioequivalence. In example studies I and II, similar bioequivalent results could have been obtained from a standard two period cross-over design; however, 160 subjects would have been enrolled in each study to ensure equivalent probability of success. Sample size can be reduced by up to 50% by using a replicate design (while the number of doses stays the same as in a typical two-period cross-over study and study duration doubles).

However, in the design of bioequivalence studies, practitioners should carefully consider the nature of their drug product at the design stage. In particular, the assumption that inter-subject variance is homogeneous (i.e. that variation associated with subject-by-formulation is null - see Hauck et al. (2000) and Zariffa et al. (2000) for more details) should be carefully considered as well as the assumption that relative bioavailability of the drug products is truly unity. Power to demonstrate bioequivalence using a standard two-period cross-over design will be reduced in either case if the assumption is violated. Allowance for these factors should be made regardless of the design employed.

It is generally the case that subject-by-formulation interaction variance may be assumed to be null when the study is planned. In this context, only roughly half the sample size required for a two-period cross-over trial need be recruited while the number of assessments remains the same. In cases where this variance is not null, the replicate design offers the additional benefit that this variance term may be separated from intra-subject variation leading to enhanced understanding of the study outcome (Grahnen et al., 1984). In a two-period cross-over, this variance is confounded with intra-subject variation leading potentially to an inconclusive study without identification of cause.

In general, studies of this nature compete for clinical resources (i.e. beds and clinic space, laboratory resources) with similar studies of brief duration. As such, it is generally of interest in situations with limited resources to limit sample size in favor of extended study duration using replicate designs. Moreover, in a statistical sense, the quality of the data is improved upon replication of administration within subject and leads to a definitive study outcome. The alternative - repeating a failed two-period cross-over study - generally leads to increased costs

beyond those incurred by using a replicate design. Of particular note, it has not been observed in published data sets that increased duration of study leads to substantially more drop-outs than the two-period cross-over design in these type of studies (Zariffa et al., 2000).

Group-sequential study designs can result in substantial cost savings when applied to bioequivalence studies. Study III was terminated at the first look, resulting in cost savings of approximately 45% relative to the full study design. Had the study been designed in a more traditional fashion (two period cross-over with no interim looks), sample size would have been as large as $n = 300$, which some might find prohibitive for demonstration of bioequivalence in a properly powered, well-controlled setting. Practitioners using such group-sequential studies should carefully consider the choice of Type I error spending function and decision rule for stopping the trial at each look when the trial is designed to ensure that adequate power is available to meet the study objectives, to ensure a clear decision on bioequivalence. Practical considerations, such as the number of beds and other resources available (sampling kits, personnel, etc.) should also be carefully considered in planning designs for highly variable products.

Study conditions and clinical procedures should be carefully monitored to ensure that homogeneity is observed across cohorts in a group-sequential study. Failure to do so may lead to incongruent variance estimates and complicate data interpretation. Under the assumption that study conditions are controlled across study parts and the study population is homogeneous, variance estimates should be homogeneous across study parts and methods of pooling variance estimates across cohorts may be applied in straightforward manner (Jennison and Turnbull, 2000). While testing procedures can be constructed to compare across cohorts, based on standard procedures for comparison of variances, it has been found that the precision of variance estimates is poor in such designs (Zariffa et al., 2000), and resulting testing procedures should be viewed with caution.

Simulations (Gould, 1995) indicate that consideration of the decision rule in conjunction with the choice of sample size when performing the first interim analysis allows for use of a slightly less conservative Type I error spending function than that proposed by Pocock (1977). Use of such a procedure results in an overall Type I error rate of precisely 5% and allows for the construction of confidence intervals as small as 93.3% for some study designs with an interim and final look. However, due to the sensitivity of such Type I error rates to logistic difficulties

encountered during the study (i.e. dropouts), it is recommended that the simple, straightforward Pocock or Bonferroni spending functions be used in practice.

If group sequential studies such as Study III continue to the second look, the variance estimate and the estimates for the mean effects of each formulation in the final analysis should be adjusted to avoid bias by the interim analysis in accordance with the procedures described by Jennison and Turnbull (2000). Practitioners designing such studies should take this adjustment into consideration when powering their group-sequential studies.

In combination, replicate and group-sequential designs make it feasible to meet and beat seemingly impossible regulatory hurdles with regard to demonstration of average bioequivalence.

Retrospective analysis of 51 replicate design data sets revealed that inference for average bioequivalence was relatively insensitive to choice of method-of-moment or REML estimation procedure. Discrepancies were observed between procedures in only a very limited subset of the data. However, inference in such situations can be problematic.

Differences in choice of model parameterisation did have an effect on model estimates. However, it was found that values of the $log-$likelihood function, and information-statistics based upon them, did not appear strikingly different between models. Where the UN model performed better, in a qualitative sense, the estimates of σ_D^2 were frequently negative, greatly complicating interpretation. These findings in combination suggest that the models' performance was not distinguishable in practice but that care should be exercised in the choice of best model based on a review of the obtained estimates.

FA0(2), CSH, and RIS appeared to be conservative procedures in that variance estimates were constrained to be null or greater, and so resulted in potentially wider confidence intervals than the corresponding Unstructured REML or method-of-moment procedures. However, this affected inference in only one, problematic data set where seemingly minor changes in estimation for variance components contributed to the findings. In the other 50 data sets, no clear discrepancy in inference was evident. It should be noted that characterisation of these variance components is also of immediate concern in PBE and IBE assessment.

Confidence intervals for ABE assessment generated using the CSH procedure will in general be wider than those generated using the FA0(2) analysis regardless of the fact that the information statistics (AIC, SBC) are the same.

Statisticians using the replicate design to establish average bioequivalence should carefully consider the type of estimation procedure to be used while designing the study. Should REML estimation be used (as would be expected in most cases where drop-outs and missing data cannot be ruled out), a procedure should be chosen which will consistently predict the outcome. If σ_D^2 estimates are expected to be non-null, it is evident that any of the four REML procedures described are roughly equivalent. However, interpretation when estimates for this variance component are null, or in unstructured models are negative, is somewhat problematic.

While the REML models appeared to perform equally well in practical application, estimates from some models may not be 'meaningful' - in the classical sense of true variances being positive. Moreover, information statistics are misleading in that equivalent values do not imply equivalent estimates of variation due to the constraint on various parameters in the likelihood.

In cases where resource constraints require that a minimal sample size be used in these studies, it is recommended that simulation studies be used in the study planning (see Chapter 5) to ensure that the choice of estimation procedure will not effect inference. Convergence of the REML estimation procedure should be carefully considered, and it may be useful to use the bootstrap to characterise the findings from the estimation procedure.

It is concluded that replicate designs may used easily and effectively to demonstrate average bioequivalence for highly variable drug products. Statisticians should, however, exercise caution in the choice of modelling procedure when using $SAS^{©}$-based approaches to the modelling of pharmacokinetic data in bioequivalence studies.

However, unanswered questions remain:

1. Do estimates from the REML models and Method-of-Moments behave as normal variables in small samples and when there is missing data?

2. What is the bias in REML estimates for the components of interest in small samples and when there is missing data?

3. Of those procedures found to provide acceptable estimates for the moments of interest, what is the Type I error rate for average bioequivalence using REML procedures in small samples?

Simulations will be conducted in Chapter 5 to address these and other questions arising from the use of replicate designs in the assessment of PBE and IBE. We now turn to these topics.

3 Small and Large Sample Properties, Estimation, and Inference for Individual Bioequivalence

The findings of this chapter were presented at the annual American Society of Clinical Pharmacology and Therapeutics meeting (Patterson et al., 2000a), at the American Association of Pharmaceutical Scientists joint workshop with the USA Food and Drug Administration (Zariffa and Patterson, 2000), at the American Statistical Society Joint Statistical meetings (Patterson and Jones, 2002e), and at the International Society of Clinical Biostatistics meeting (Patterson and Jones, 2002f). Aspects of the findings were published in the *Journal of Clinical Pharmacology* (Zariffa and Patterson, 2001), in *Pharmaceutical Statistics* (Patterson and Jones, 2002a; 2002g), in the *Proceedings of the Joint Statistical Meetings* (Patterson and Jones, 2002h), and in a series of a GlaxoSmithKline technical reports (Patterson et al., 2001e; Patterson and Jones, 2002b-c). Aspects of the findings relating to use of an alternative individual bioequivalence metric, the Kullback-Leibler distance, were published in Dragalin et al. (2002) but will not be discussed further in this thesis.

3.1 Introduction, Previous Research, and Goals of Chapter

In this Chapter, key ideas in individual bioequivalence will be quickly reviewed, and previous research by the author will be summarised in order to illustrate the genesis of research topics explored in the remainder of the Chapter.

3.1.1 Review

We now turn to detailed discussion of the use of retrospective analysis and simulation in bioequivalence assessment. Following review of previous research, inferential procedures in individual (IBE) bioequivalence will be assessed, and it will be shown how retrospective analysis of data is used to assess performance of the metrics in bioequivalence assessment. Such an exercise is unlikely to yield definitive conclusions as to the proposed PBE and IBE criteria due to the usual caveats associated with relatively small retrospective analyses. It does however offer the likelihood of providing useful observations and possibly highlighting key areas where further investigations are needed. As such it deepens our understanding of the issues involved so that

a continually improved and more informed dialogue can take place. Simulations will be used in a subsequent Chapter to assess hypotheses arising from this exercise, and questions remaining to be addressed by additional data collection will be described.

To review, average bioequivalence (ABE; FDA Guidance, 1992) has traditionally been used as the standard for market access with regulatory limits of twenty percent. This approach was discussed in Chapter 2 and will not be discussed further here.

In bioequivalence studies, using a replicate design with sequences RTRT and TRTR, the following mixed model for log_e-transformed observations is commonly accepted (Jones and Kenward, 1989). Let Y_{tjk} be the k-th response ($k = 1, 2, \ldots$) for the j-th subject in the cross-over trial administered formulation t ($t = T, R$) and

$$Y_{tjk} = \xi_{tj} + \varepsilon_{tjk} = \mu_t + \nu_{tj} + \varepsilon_{tjk}$$

where

ν_{tj} and ε_{tjk} are independent with mean zero

$Var(\nu_{tj}) = \sigma^2_{Bt}$, the between-subject variance,

$Var(\nu_{Tj} - \nu_{Rj}) = \sigma^2_D$, the subject-by-formulation interaction variance,

$$Cov(\nu_{Tj}, \nu_{Rj}) = \rho\sigma_{BT}\sigma_{BR},$$

$Var(\varepsilon_{tjk}) = \sigma^2_{Wt}$, the within-subject variance,

$$Cov(\varepsilon_{tjk}, \varepsilon_{tjk'}) = 0, \text{ for } k \neq k'.$$

Note that nuisance effects (period and sequence effects) may be fitted in practice (Jones and Kenward, Chapter 4, 1989) but are omitted from the above description for the sake of clarity. The above model may be fitted using general linear models (corresponding to a method-of-moments approach), maximum likelihood or restricted maximum likelihood based procedures. This type of study design has been used for some time to assess average bioequivalence (Chapter 2) but was proposed for a different purpose in FDA Guidance (1997).

Following the original proposal in 1997 (cf. FDA Guidance), in August 1999, the FDA re-proposed new guidelines for the assessment of bioequivalence: population bioequivalence (PBE) and individual bioequivalence (IBE) (FDA Guidances, 1999a and 1999b) and finalised proce-

dures in 2000-2001 (FDA Guidance, 2000b, 2001) based on ideas developed by Anderson (1993) and Anderson and Hauck (1983, 1990). In the case of pre-market approval, one can formulate the bioequivalence question as "Can a patient begin their therapy with either formulation (commercial or clinical trial) and be assured same results in terms of safety and efficacy?" This has been called the concept of prescribability (Anderson and Hauck, 1990) and is linked to population bioequivalence (PBE). This topic will be addressed in Chapter 4.

In the case of post marketing changes, the bioequivalence question becomes: "Can I safely and effectively switch my patient from their current formulation to another?" This has been called the concept of switchability (Anderson and Hauck, 1990) and is linked to IBE. The criteria used to assess IBE under the proposed FDA draft guidance (1997-1999) and finalised procedures in 2000-2001 (FDA Guidance, 2000b, 2001) aggregates the difference between population means and variances and accounts for subject predictability from one formulation to the other (subject-by-formulation interaction, Ekbohm and Melander, 1989). In addition, the individual bioequivalence metric allows for scaling of the regulatory limits based on the within-subject variability of the reference product.

Individual bioequivalence is assessed using the following aggregate statistic (FDA Guidance, 1997).

$$\frac{(\mu_T - \mu_R)^2 + \sigma_D^2 + \sigma_{WT}^2 - \sigma_{WR}^2}{max(0.04, \sigma_{WR}^2)}$$

Because the within-subject variance of each formulation cannot be separately estimated from between-subject variance estimates in most two-period cross-over designs of the form { TR, RT }, a replicate design is generally required. It should be noted that if the Huyhn-Feldt condition (1970) is assumed for between-subject variability across formulations, it is sometimes possible to estimate within-subject variability for each formulation using a restricted or maximum likelihood approach to inference. However, under this approach, subject-by-formulation interaction, σ_D^2, is assumed to be null. Because the within-subject variance of each treatment can not usually be reliably and separately estimated in a two-way cross-over { TR, RT } (unless the assumption that between-subject variability is homogeneous is made along with the assumption that correlation

is unity), a replicated design (a cross-over with sequences { TRTR, RTRT }) is required for IBE assessment.

Note that, due to the nature of this 'aggregate' criterion, differences in means in this criterion can be 'negated' by decreased within-subject variance for the test formulation. Some have noted this to be an undesirable property of the proposed metric, (Endrenyi and Hao, 1998c), and it is known that such trade-offs do occur in practice (Zariffa et al., 2000). Additionally, it should be noted that the between-subject inconsistency of the test-reference comparison quantified by estimated σ_D^2 might be an inadequate measure of the switchability (Zariffa et al., 2000).

3.1.2 Previous Research

This retrospective analysis (published in Zariffa et al., 2000 and performed by the author) will now be discussed in more detail. Analysis was performed on 22 data sets from 15 replicate cross-over bioequivalence studies (see Data sets A through O2, Section 2.7.1) and is summarized below. AUC and Cmax parameters from these studies were analyzed using average, population, and individual bioequivalence methods (FDA Guidance, 1997). Practical issues involving the behavior of the new criteria and its expected impact on sample size for highly variable drug products were presented, and the characterization of key parameters and their inter-relationships were discussed with particular emphasis on the subject by formulation term in the individual bioequivalence criteria. It was concluded more studies and simulations were desirable before full-scale implementation of population and individual bioequivalence criteria.

In each study, subjects provided data on four separate sessions separated by adequate washout to avoid residual drug concentrations from the previous occasion as described in Chapter 1. Summary measures AUC and Cmax were derived in each period (as six of the 15 studies involved multiple drug components, a total of 22 data sets were available).

These data were subjected to statistical analysis under the average bioequivalence guidance from FDA (cf. FDA Guidance, 1992) as follows. Log_e-transformed AUC and Cmax were modelled separately using a general linear model (GLM in $SAS^{©}$) with terms accounting for sequence, subject within sequence, period, formulation, and formulation by subject within sequence interaction. The contrast $\hat{\mu}_T - \hat{\mu}_R$ (where T=test and R=reference formulation) was derived based on the difference in least squares means, and a 90% confidence interval is derived

based on the mean-squared error estimate for the formulation by subject within sequence interaction term. These quantities were then exponentiated to derive the ratio of geometric means (PE= $e^{\hat{\mu}_T - \hat{\mu}_R}$) and a corresponding confidence interval. If this confidence interval falls completely within the interval $(0.80 - 1.25)$ for both AUC and Cmax, then average bioequivalence was demonstrated.

Under the proposed guidance for population and individual bioequivalence (FDA Guidance, 1997), analyses were conducted using a two stage, restricted maximum likelihood model (model (22)) where estimates $\hat{\mu}_T$, $\hat{\mu}_R$, $\hat{\sigma}^2_{BT}$, $\hat{\sigma}^2_{BR}$, $\hat{\rho}$, $\hat{\sigma}^2_D$, $\hat{\sigma}^2_{WT}$, $\hat{\sigma}^2_{WR}$ are derived as appropriate to the criteria under study ((23) for population bioequivalence and (24) for individual bioequivalence). Inference was assessed based on 2000 bootstraps (Efron and Tibshirani, 1993). If the upper ninety-five percent bound on the FDA metric, based on the nonparametric-percentile method (Efron and Tibshirani, 1993), falls below the value of 1.74 for population bioequivalence and 2.49 for individual bioequivalence for both AUC and Cmax, then bioequivalence is demonstrated for the endpoint under study as appropriate. See Section 1.5 for discussion of derivation of these goalposts. Analyses were performed on COMPAQ Deskpro Pentium machines using *SAS*© for Windows 6.11, and an external statistician validated results on a subset of five data sets.

Of the 22 data sets for AUC, 19 pass average bioequivalence, all pass PBE and 20 pass IBE (see Figure 15). Of the three data sets that failed average bioequivalence, all passed PBE and one passed IBE. The results for Cmax are more variable (see Figure 16). Of the 16 data sets where average bioequivalence is demonstrated, one data set failed both PBE and IBE. Of the six data sets that failed average bioequivalence, two passed both PBE and IBE, three passed PBE but not IBE and one failed all three criteria. There were five data sets that passed average bioequivalence and PBE but not IBE.

Image removed due to third party copyright

Figure 15: Results of retrospective analyses of AUC in SB database of 22 data sets: Average, Population, and Individual Bioequivalence (Zariffa et al., 2000)

Image removed due to third party copyright

Figure 16: Results of retrospective analyses of Cmax in SB database of 22 data sets: Average, Population, and Individual Bioequivalence (Zariffa et al., 2000)

As described in Chapter 1, the proposed criteria for PBE and IBE are based on aggregate test statistics which tradeoff differences in means relative to difference in variances. Data set N1 for Cmax (Figure 16) is an example where a modest mean difference of 6% is overshadowed by an increase in the test formulation variability for both PBE and IBE. The IBE criteria further penalizes the sponsor of a bioequivalence trial when the estimate of σ_D is non-zero. Data set M for Cmax (Figure 16) is an example of this feature. Here the mean and variance differences are negligible but the estimate for subject-by-formulation interaction standard deviation is quite large (0.23) contributing directly to the failure to demonstrate IBE. Last, the effect of scaling to the reference formulation in PBE and IBE can offset both differences in means and increases in the test formulation variability. As an example of this, consider data set I2 (Figure 15) where a 33% reduction in mean AUC is offset by the scaling in PBE. This same data set could not pass IBE due to a large subject by formulation interaction. These few examples from the database highlight the complexities and inter-relation between the various components in the PBE and IBE criteria.

As the exact distribution of the test statistics for both PBE and IBE were not known at the time the draft guidance was issued (FDA Guidance, 1997), a bootstrap procedure was suggested by FDA for inference assessment. The bootstrap procedure gives rise to curious phenomena in the boundary region when data sets exhibit reference product variances just above the cutoff value of 0.04. In such cases, the decision to reference or constant scale is made based on the original data set and each bootstrap sample is then scaled independently to the bootstrap sample's or the constant value of σ_R^2 for PBE and of σ_{WR}^2 for IBE. Given the distribution of the variance component, the aggregate test statistic from the separate bootstrap samples may be over estimated as the denominator term is often lower than the value in the original data set and the cutoff of 0.04. A striking example of this feature is provided by the Cmax data in data set E (Figure 16). Here the estimate of σ_{WR} is 0.204, just above the standard deviation cutoff value of 0.20. Using the reference scaled procedure as outlined in the draft guidance, we achieve an upper 95th percentile of 3.67 for IBE thus failing to demonstrate IBE. It should be noted that using a constant scaled IBE approach (which is theoretically more conservative), the upper 95th percentile of the bootstrap distribution is 2.08 indicating data set E would have passed IBE.

It was anticipated that minor modifications to the bootstrap procedure could remedy this curious situation. Alternatively, sponsors could be allowed to make the decision to reference or constant scale based on the properties of the specific data set for the study being analyzed.

Reports subsequently published separately (Shao et al., 2000a-b) indicated that the bootstrap procedure proposed by FDA (cf. 1997 Guidance) is consistent when estimates for the reference product variation differ from 0.04 and discussed alterations of the bootstrap procedure to ensure accurate, precise estimates in the neighborhood of 0.04. Conservation of Type I error using these approaches however was not well studied (we will address this in Chapter 5), and we will consider practical performance of this approach using the database of replicate cross-over designs.

It should be noted that sponsors of bioequivalence studies would bear additional resource burden for products with low to moderate variability in the reference formulation to assess IBE (see Table 14 below). In contrast, the scaling feature of the PBE and IBE criteria lead to a decrease in sample size requirements for highly variable drug products (Table 14). In order to demonstrate this, random selections of data from half the subjects in each sequence in data set B (see Section 2.7.1) were taken, and there was no difficulty in demonstrating PBE and IBE (data not shown).

Table 14: Sample Sizes for Average and Individual Bioequivalence in a Four-period, Replicate Cross-over Design (FDA Draft Guidance, 1999b)

σ_{WR}	N for ABE	N for IBE
0.15	8	14
0.23	16	30
0.30	28	36
0.50	72	36
Assumptions: 90% power, $\sigma_D = 0.01$, $\sigma_{WT} = \sigma_{WR}$		

Even without further data and studies, it was clear that at least some features of the PBE and IBE were undesirable. Chief was the use of the bootstrap procedure as suggested (FDA Guidance, 1997), particularly, the peculiar behavior of the methodology in the boundary region of 0.04 for scaling to reference product variation. This problem for scaling to reference variation would not be easily resolved in practice and had the potential to lead industry sponsors and regulators into unreasonable discussions regarding the final inference in such data sets.

Finally, it should be noted that the scaling of the regulatory limits for compounds with

intrinsically large variance is a welcome feature of the proposed criteria, but this can also achieved within the average bioequivalence framework, for example by using scaled-average bioequivalence (Midha et al., 1997a-b).

While some understanding can be gained by the above individual review of the various data sets and the reasons leading to the changes in inference under each criteria, careful examination of the various components of the criteria and their inter-dependencies across the entire database was more revealing. Such an approach was used in evaluating the key term of subject by formulation (σ_D) in the IBE criteria. See Chapter 1 for discussion of the importance of this parameter in IBE assessment.

Subject predictability, as expressed by σ_D, is of particular concern in assessing switchability. In three of 22 data sets, estimated AUC σ_D exceed the 0.15 value. For Cmax, a more variable endpoint, the frequency of estimates exceeding the 0.15 cutoff was eight of 22 in this database. This might suggest a possible relationship between estimates of σ_D and the intrinsic variance of the PK metric itself. Further, when considering the 90% bootstrap CI for estimated σ_D, the value of 0.15 was often contained in the confidence interval. In fact, this occurred in nine and 19 data sets for AUC and Cmax respectively. These large estimates of σ_D were associated with failure to demonstrate IBE in one of three cases for AUC and six of eight cases for Cmax (Figures 15 and 16). However, care needs to be taken when evaluating the magnitude of σ_D in isolation of other parameters of interest. Indeed, when reviewing the relationship between the subject predictability term and the inherent variation in the estimates for σ_{WR} a possible positive correlation was observed (Figure 17). Again, it is noted that the SB database did not contain many examples of high estimated σ_D coupled with low σ_{WR}. This can be due to a true positive relationship between the two parameters or, equally possible, an observation bias in this database.

Figure 18 offers the characterization of estimated σ_D as a function of estimates for the difference in between-subject test and reference standard deviations ($\sigma_{BT} - \sigma_{BR}$) and twice the product of the correlation and between-subject standard deviations ($2(1-\rho)\sigma_{BT}\sigma_{BR}$). Here it is clear that there is a relationship between estimates for σ_D and the absolute difference between σ_{BT} and σ_{BR}. More curious are the large number of data sets with estimates for $2(1-\rho)\sigma_{BT}\sigma_{BR}$ of zero, associated with an estimated correlation of unity. Indeed, half of all data

sets exhibited estimates for correlation of unity using the PROC MIXED procedure described in the guidance. This observation deserves further consideration especially as alternative mixed modelling procedures and parametric inferential methods are entertained (and will be discussed further in this Chapter.) As σ_D is a key ingredient in the IBE criteria, these features were deserving further study, through the use of simulation, as will be considered in Chapter 5.

Image removed due to third party copyright

Figure 17: Estimated Subject by Formulation Interaction (σ_D) and 90% bootstrap CI versus Estimated Within-subject Reference Product Standard Deviation (σ_{WR}) for AUC and Cmax (Zariffa et al., 2000)

Image removed due to third party copyright

Figure 18: Estimated Subject by Formulation Interaction Standard Deviation (σ_D) in Relation to the Difference in Between-Subject Variability in Test and Reference Products ($\sigma_{BT} - \sigma_{BR}$) and the Corrected Correlation ($2(1-\rho)\sigma_{BT}\sigma_{BR}$) (Zariffa et al., 2000)

Further observations regarding σ_D can be found in the work of Endrenyi et al. (2000). In this paper it was established that the method-of-moments estimator $\hat{\sigma}_D^2 = \hat{\sigma}_I^2 - (\frac{\hat{\sigma}_{WT}^2 + \hat{\sigma}_{WR}^2}{2})$ is unbiased (Chinchilli and Esinhart, 1996) for σ_D^2 with variance

$$2\left[\frac{(\frac{\sigma_{WT}^2+\sigma_{WR}^2}{2} + \sigma_D^2)^2}{n-1} + \frac{(\frac{\sigma_{WT}^2+\sigma_{WR}^2}{2})^2}{n}\right]$$

in balanced, replicate designs.

We now turn to considerations for further research. While there is little to no evidence to suggest that the ABE criteria has failed to protect the public (Barrett et al., 2000), consumers may find the specific questions above related to PBE and IBE more relevant. One reviewer noted recently that very little data have been published to assess how the proposed FDA criterion perform (Colburn and Keefe, 2000). There is some level of intuitive appeal for both population and individual bioequivalence although it is expected the individual criteria would be more relevant to consumers as they 'switch' to generic formulations of chronically administered medications.

It should be noted here that one of the key provisions of the 1999 draft FDA guidance was the suggestion for what has been termed a 'public health experiment' or mandatory data collection period (Montreal, AAPS/FDA Workshop, August/September 1999). Sponsors of any bioequivalence study would be compelled to submit data from a replicate design to FDA for approval to market. Subsequent discussion at the Advisory Committee for Pharmaceutical Science (September 1999, Washington DC) resulted in the recommendation that market access not be permitted unless average bioequivalence had been demonstrated under existing criteria. Restricting the 'experiment' to a class of drugs such as controlled release formulation and highly variable drugs was also suggested. These proposals were adopted in the finalised guidance (FDA Guidance, 2000b) but were subsequently removed (FDA Guidance, 2002).

This calls for careful and meticulous examination of existing replicate design data sets prior to concluding the 'public health experiment' so as to set provide a comprehensive outcome for the exercise and for the careful, scientific consideration of viable alternatives to the procedure developed by the FDA.

It should be noted that ideas relating to assessment of population bioequivalence have been underdeveloped in the statistical literature. Attention has focussed on individual bioequivalence in the majority of publications and meetings on the topic (see Chapter 1). Note that the FDA

metric (23) does not account for each subject as their own control as is the case in cross-over designs. Ignoring this relationship is a classical error in cross-over designs (Jones and Kenward, 1989; Senn, 1993; Senn, 2002). A framework proper to the assessment of population bioequivalence in cross-over designs will be developed in Chapter 4.

3.1.3 Goals of Chapter

In this chapter we will develop frequentist procedures for assessment of individual bioequivalnce. Emphasis will be placed upon the use of unbiased (or asymptotically unbiased) procedures appropriate to the study design performed and data collected. Retrospective analysis will be used to assess the performance of procedures developed in practice, and simulation studies to assess any hypotheses resulting from the exercise will be defined for Chapter 5.

Estimation and inferential procedures for FDA proposed assessment in individual bioequivalence assessment will be explored, and an alternative testing procedure to those currently available will then be developed.

Retrospective analysis using the procedures developed in the previous Chapters and this one will be conducted and discussed based on the database of replicate cross-over designs described in Section 2.7.1. Based on these findings, a plan and program for simulation studies are prepared to assess hypotheses generated from the retrospective analyses. Discussion of the findings and issues to be addressed in the remainder of the thesis will conclude discussion in this Chapter.

We adopt the Method-of-Moment and REML procedures described in Chapter 2 for the estimation of moments involved in individual bioequivalence assessment. See Chapter 2 for further details.

3.2 Properties of the Estimated Metrics for IBE Assessment

Based on the above findings, we will first show that the metric proposed by the FDA in complete data sets (FDA Guidance, 1997, 1999a, 1999b, 2000b, 2001) for the assessment of individual bioequivalence is asymptotically unbiased, though in small samples it carries a small positive bias related to degrees of freedom using method-of-moments estimation. We will then turn to the study of its properties using REML models described previously.

The FDA (2001) guidance specifies that both the constant and reference scaled metrics

should be constructed independent of the level of reference product variation and the approach to estimation developed in this thesis were studied under this approach. Hence the expectations and variances are derived independent of the level of estimated within-subject reference variation.

Within-subject reference variation is subsequently used to determine which is the most appropriate criteria to look at, but not necessarily to post hoc determine which one is the most appropriate for use. Indeed FDA guidance suggests using both constant and reference scaled metrics under certain circumstances, ' VII. D. Discontinuity The mixed-scaling approach has a discontinuity at the changeover point, sW0 (individual BE criterion) or sT0 (population BE criterion), from constant- to reference-scaling. For example, if the estimate of the within-subject standard deviation of the reference is just above the changeover point, the confidence interval will be wider than just below. In this context, the confidence interval could pass the predetermined BE limit if the estimate is just below the boundary and could fail if just above. This guidance recommends that sponsors applying the individual BE approach may use either reference-scaling or constant-scaling at either side of the changeover point. With this approach, the multiple testing inflates the type I error rate slightly, to approximately 6.5%, but only over a small interval of sWR (about 0.18-0.20).'.

In all cases we will first investigate reference-scaled metrics followed with discussion of constant-scaled metrics.

Theorem 3.1 *Bias in reference-scaled FDA Metric Estimated using Method-of-Moments*

The reference-scaled FDA metric (24) may be estimated (FDA Guidance, 1999a, 1999b, 2000b, 2001) using the method of moments approach as:

$$\frac{\hat{\delta}^2 + M_I + \frac{M_T}{2} - \frac{3M_R}{2}}{M_R} = \frac{\hat{\delta}^2}{M_R} + \frac{M_I}{M_R} + \left(\frac{1}{2}\right)\left(\frac{M_T}{M_R}\right) - \frac{3}{2} \tag{33}$$

Note that M_I is an unbiased estimate for $\sigma_I^2 = \sigma_D^2 + \frac{\sigma_{WT}^2 + \sigma_{WR}^2}{2}$.

This statistic is a consistent estimator and furthermore is asymptotically unbiased and positively biased in small samples with expected value

$$\frac{\nu}{\nu - 2}\left(\frac{\delta^2}{\sigma_{WR}^2} + \frac{(n+1)\sigma_I^2}{n\sigma_{WR}^2} + \frac{\sigma_{WT}^2}{2\sigma_{WR}^2}\right) - \frac{3}{2} \tag{34}$$

Proof: Let (27) with corresponding assumptions hold. Then it follows directly that $\frac{\hat{\delta}^2}{\sigma_I^2/n} \sim \chi_1^{2'}(\frac{\delta^2}{\sigma_I^2/n})$ (Vonesh and Chinchilli, 1997) where $\chi_1^{2'}$ represents a non-central chi-squared distribution with one degree of freedom and non-centrality parameter $\frac{\delta^2}{\sigma_I^2/n}$ (Muirhead, p22, 1982). As (33) is an estimate for (24) obtained from method of moments and (24) is continuous, then (33) is consistent (Bickel and Doksum, 1977).

Taking the expectation of terms in (33) and assuming terms are pairwise independent under Theorem 2.1,

$$E\Big(\frac{\hat{\delta}^2}{M_R} + \frac{M_I}{M_R} + \Big(\frac{1}{2}\Big)\Big(\frac{M_T}{M_R}\Big) - \frac{3}{2}\Big)$$

$$= E\Big(\frac{\frac{\sigma_I^2}{n}(\frac{\hat{\delta}^2}{\sigma_I^2/n})}{\sigma_{WR}^2(\frac{M_R}{\sigma_{WR}^2})}\Big) + E\Big(\frac{\sigma_I^2(\frac{M_I}{\sigma_I^2})}{\sigma_{WR}^2(\frac{M_R}{\sigma_{WR}^2})}\Big) + E\Big(\frac{\sigma_{WT}^2(\frac{M_T}{\sigma_{WT}^2})}{2\sigma_{WR}^2(\frac{M_R}{\sigma_{WR}^2})}\Big) - \frac{3}{2}$$

Further, using the results of Muirhead (p 24, 1982), it is seen that this expression reduces to,

$$E\Big(\frac{\sigma_I^2}{n\sigma_{WR}^2}F_1'\Big) + E\Big(\frac{\sigma_I^2}{\sigma_{WR}^2}F_1\Big) + E\Big(\frac{\sigma_{WT}^2}{2\sigma_{WR}^2}F_2\Big) - \frac{3}{2}$$

where F_i' is a random variable with non-central $F_{1,\nu}'(\frac{\delta^2}{\sigma_I^2/n})$-distribution with non-centrality parameter $\frac{\delta^2}{\sigma_I^2/n}$ and $1, \nu$ degrees of freedom. F_i is a random variable distributed according to the central-$F_{\nu,\nu}$-distribution with ν, ν degrees of freedom. Taking the expectation (Muirhead, p 25 1982), we see that the result is (34). As sample size increases,

$$lim_{n\to\infty}\Big[\frac{\nu}{\nu-2}\Big(\frac{\delta^2}{\sigma_{WR}^2} + \frac{(n+1)\sigma_I^2}{n\sigma_{WR}^2} + \frac{\sigma_{WT}^2}{2\sigma_{WR}^2}\Big) - \frac{3}{2}\Big] = \frac{\delta^2}{\sigma_{WR}^2} + \frac{\sigma_I^2}{\sigma_{WR}^2} + \frac{\sigma_{WT}^2}{2\sigma_{WR}^2} - \frac{3}{2}$$

which is an unbiased estimate for (24). However in small samples, the bias is

$$\Big(\frac{\nu}{\nu-2} - 1\Big)\frac{\delta^2}{\sigma_{WR}^2} + \Big(\frac{\nu(n+1)}{n(\nu-2)} - 1\Big)\frac{\sigma_I^2}{\sigma_{WR}^2} + \Big(\frac{\nu}{\nu-2} - 1\Big)\frac{\sigma_{WT}^2}{2\sigma_{WR}^2} \geq 0$$

□□□

Theorem 3.2 *Bias in constant-scaled FDA Metric Estimated using Method-of-Moments*

The constant-scaled FDA metric (24) may be estimated (FDA Guidance, 1999a, 1999b, 2000b,

2001) using the method of moments approach as:

$$\frac{\hat{\delta}^2 + M_I + \frac{M_T}{2} - \frac{3M_R}{2}}{0.04} = \frac{1}{0.04}(\hat{\delta}^2 + M_I + \left(\frac{1}{2}\right)\left(M_T\right) - \frac{3M_R}{2}) \tag{35}$$

and has expected value

$$\frac{\frac{\sigma_I^2}{n} + \delta^2 + \sigma_D^2 + \sigma_{WT}^2 - \sigma_{WR}^2}{0.04}$$

Proof: Under the approach developed in Theorem 3.1, consider (35). Taking the expectation of each term (Muirhead, p24-27, 1982), it is found that this is

$$\frac{1}{0.04}\left(\frac{\sigma_I^2}{n} + \delta^2 + \sigma_I^2 + \frac{\sigma_{WT}^2}{2} - \frac{3\sigma_{WR}^2}{2}\right)$$

$$= \frac{\frac{\sigma_I^2}{n} + \delta^2 + \sigma_D^2 + \sigma_{WT}^2 - \sigma_{WR}^2}{0.04}$$

□□□

Thus the estimation procedure is positively biased (against sponsors) when using the recommended method-of-moments estimation procedure (FDA Guidance, 1999a, 1999b, 2000b, 2001). An unbiased estimator for the metric is now provided.

Theorem 3.3 *Unbiased FDA Metrics for IBE Estimated using Method-of-Moments*

The reference-scaled FDA metric (24) may be estimated in an unbiased fashion in small samples using the method of moments approach as:

$$\frac{\nu - 2}{\nu}\left[\frac{\hat{\delta}^2}{M_R} + \frac{(n-1)M_I}{nM_R} + \left(\frac{1}{2}\right)\left(\frac{M_T}{M_R}\right)\right] - \frac{3}{2} \tag{36}$$

A constant-scaled unbiased estimator for the FDA metric is:

$$\frac{\hat{\delta}^2 + (1 - \frac{1}{n})M_I + \frac{M_T}{2} - \frac{3M_R}{2}}{0.04}$$

Proof: Let (27) with corresponding assumptions hold. Taking the expectation of terms in (36) and assuming terms are pairwise independent under Theorem 2.1 and using the results of

Muirhead (p 24, 1982), it is seen that this expression reduces to,

$$E\Big(\frac{(\nu-2)\sigma_I^2}{\nu(n\sigma_{WR}^2)}F_1'\Big)+E\Big(\frac{(\nu-2)(n-1)\sigma_I^2}{n\nu\sigma_{WR}^2}F_1\Big)+E\Big(\frac{(\nu-2)\sigma_{WT}^2}{2\nu\sigma_{WR}^2}F_2\Big)-\frac{3}{2}$$

where F_i' and F_i are random variables with the non-central and central-F-distributions as previously. Taking the expectation (Muirhead, p 25 1982), we see that the result is

$$\frac{\delta^2}{\sigma_{WR}^2}+\frac{\sigma_I^2}{\sigma_{WR}^2}+\frac{\sigma_{WT}^2}{2\sigma_{WR}^2}-\frac{3}{2}$$

which is unbiased for the FDA metric (24).

Similarly, taking the expectation of $\frac{\hat{\delta}^2+(1-\frac{1}{n})M_I+\frac{M_T}{2}-\frac{3M_R}{2}}{0.04}$, it is found (Muirhead, p24-27, 1982) that this expression has expectation:

$$\frac{\frac{\sigma_I^2}{n}(1+\frac{\delta^2}{\sigma_I^2/n})+(1-\frac{1}{n})\sigma_I^2+\frac{\sigma_{WT}^2}{2}-\frac{3\sigma_{WR}^2}{2}}{0.04}$$

□□□

Estimates from a REML model may be utilised in construction of the metric of interest in similar fashion to that used for the method-of-moments procedure to estimate the metrics of interest. However, the estimators are known to be only asymptotically unbiased and asymptotically normally distributed based on previously established results (Searle, 1971). Here $\underline{\hat{\beta}}$ and $\hat{\boldsymbol{\Sigma}}$ are REML estimates for $\underline{\beta}$ and $\boldsymbol{\Sigma}$. Then based on Rao (1973) and as an extension to the results of Searle (1971), the large sample variance for $\underline{\hat{\beta}}$ and $\hat{\boldsymbol{\Sigma}}$ are $(\boldsymbol{X}'\boldsymbol{\Sigma}^{-}\boldsymbol{X})^{-}$ and $-E[\frac{\partial^2 \boldsymbol{L}^{-1}}{\partial\boldsymbol{\Sigma}\partial\boldsymbol{\Sigma}'}]$, respectively with covariance $\mathbf{0}$ where $\boldsymbol{L}$ is the *log*-likelihood in expression (30). We discuss application of this method in the next section.

3.3 Testing and Inferential Procedures for IBE

We first turn to discussion of the goalpost in IBE assessment. These were discussed in Chapter 1 but are briefly summarised here. The goalpost for individual bioequivalence assessment assumes a within-subject variance for the reference formulation of 0.04 and is set to 2.49 as follows

$$\frac{(log_e(1.25))^2+(0.03)+(0.02)}{0.04}$$

allowing for a mean difference of twenty percent and a variance allowance of 0.03 in the numerator for subject-by-formulation interaction and 0.02 for the difference in within-subject variance under the procedure proposed by the FDA (FDA Guidance, 1997). If the upper ninety-five percent bound on the FDA metric falls below this value of 2.49, individual bioequivalence is demonstrated for the endpoint under study.

However, from a scientific perspective given knowledge of the estimates from the database in Chapter 2, this would appear to allow for quite large changes in average exposure dependent on the magnitude of σ_D^2 and the difference in within-subject variability. Alternatively, as we know that the estimates for $\mu_T - \mu_R$ and σ_D^2 are usually near 0, the metric would appear to allow for large increases in σ_{WT}^2 dependent on the magnitude of σ_{WR}^2. The response-surface and projected contour plots describing these 'trade-offs' may be found in Figure 19.

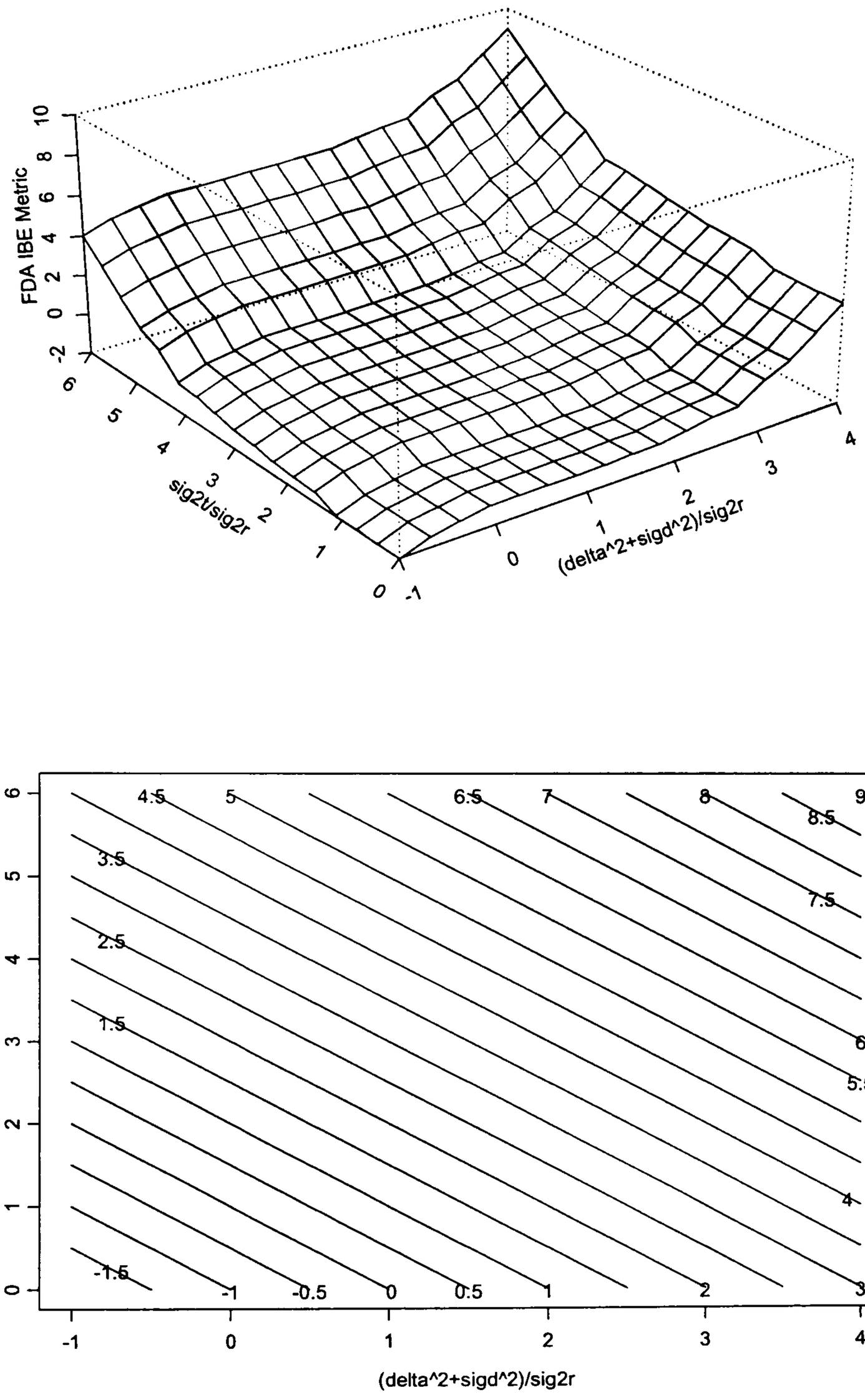


Figure 19: Response-Surface and Projected Values of FDA IBE Metric= $x + y - 1$ relative to $x = \frac{(\mu_T - \mu_R)^2 + \sigma_D^2}{\sigma_R^2}$ and $y = \frac{\sigma_T^2}{\sigma_R^2}$

However, definition of the regulatory acceptance bound is a regulatory responsibility, and while we question the approach that the FDA has adopted and recommend they reconsider based on inspection of data and plots such as Figure 19, for the purposes of the work in this thesis, we will utilise the goalpost defined in the FDA Guidance (1997, 1999b, 2001).

Note that the results of the previous section allow for straightforward application of the non-parametric percentile method (Efron and Tibshirani, 1993; Shao and Tu, 1996) to construct confidence intervals for the metric of interest. The properties of these procedures have been established separately (Shao et al., 2000a-b) but have only received limited study in practice (Zariffa et al., 2000) and have not been studied with regard to protection of Type I error. We will amend these deficiencies in this thesis. However, it is obviously desirable to have an approximate or asymptotic procedure to assess inference.

In complete data sets, inferential procedures for such the FDA metric may be based on approximation procedures such as the Cornish-Fisher expansion (Hyslop et al., 2000). Under this approach, the one-sided null hypothesis

$$H_0 : \frac{\delta^2 + \sigma_D^2 + \sigma_{WT}^2 - \sigma_{WR}^2}{max(0.04, \sigma_{WR}^2)} \geq c_{FDA} \tag{37}$$

is tested where $c_{FDA} = 2.49$. This expression is linearised as

$$H_0 : \nu_{IBE} = \delta^2 + \sigma_D^2 + \sigma_{WT}^2 - (1 + c_{FDA})\sigma_{WR}^2 \geq 0 \tag{38}$$

. For low variability compounds the expression is linearised as:

$$H_0 : \nu_{C.IBE} = \delta^2 + \sigma_D^2 + \sigma_{WT}^2 - \sigma_{WR}^2 - 0.04(c_{FDA}) \geq 0$$

Method-of-moments estimates for the moments of interest may be placed in this expression and the Cornish-Fisher expansion (Johnson, et al., 1994) may be applied to calculate an approximate upper bound (Hyslop et al., 2000) for complete data sets using confidence bounds for each of the parameters as follows.

For the linearised version of the FDA's IBE metric, a procedure is described in the FDA Guidance (2001) based on Hyslop et al. (2000) that is appropriate for replicate cross-over de-

signs with no missing data and is summarised as follows. In this situation, the estimates $\hat{\delta}$, $\hat{\sigma}_I^2$, $\hat{\sigma}_{WT}^2$, and $\hat{\sigma}_{WR}^2$ are derived for $\delta = \mu_T - \mu_R$, $\sigma_I^2 = \sigma_D^2 + \frac{\hat{\sigma}_{WT}^2 + \hat{\sigma}_{WR}^2}{2}$, σ_{WT}^2, and σ_{WR}^2 based on method-of-moment estimates as follows. Let the individual difference across formulations be denoted $I_{ij} = \bar{y}_{Tij\bullet} - \bar{y}_{Rij\bullet}$ such that

$$\hat{\delta} = \frac{1}{s}\sum_{i=1}^{s}\left[\frac{1}{n_i}\sum_{j=1}^{n_i} I_{ij}\right]$$

and

$$M_I = \frac{1}{(\sum_{i=1}^{s} n_i) - s}\sum_{i=1}^{s}\sum_{j=1}^{n_i}(I_{ij} - \bar{I}_i)^2$$

Let the individual difference within formulations for test and reference formulations be denoted $T_{ij} = y_{Tij1} - y_{Tij2}$ and $R_{ij} = y_{Rij1} - y_{Rij2}$, respectively. Within-subject variances are estimated by

$$M_T = \frac{1}{2((\sum_{i=1}^{s} n_i) - s)}\sum_{i=1}^{s}\sum_{j=1}^{n_i}(T_{ij} - \overline{T}_i)^2$$

and

$$M_R = \frac{1}{2((\sum_{i=1}^{s} n_i) - s)}\sum_{i=1}^{s}\sum_{j=1}^{n_i}(R_{ij} - \overline{R}_i)^2$$

The estimates $\hat{\delta}$, M_I, M_R and M_T are known to be pair-wise independent (Theorem 2.1) in complete, balanced data sets.

Note that $\frac{\hat{\delta}^2}{\sigma_I^2/n} \sim \chi_1^{2'}(\frac{\delta^2}{\sigma_I^2/n})$ (Vonesh and Chinchilli, 1997) where $\chi_1^{2'}$ represents a non-central chi-squared distribution with one degree of freedom and non-centrality parameter $\frac{\delta^2}{\sigma_I^2/n}$ (Muirhead, p22, 1982). Hence, when these estimates are 'plugged-in' to the equation for the FDA metric, it is composed of a linear function of independent chi-squared variables. The following approach is then used to derive an approximate 90% confidence interval.

1. Derive unbiased, independent method-of-moments estimators $\hat{\delta}$, $\hat{\sigma}_I^2$, $\hat{\sigma}_{WT}^2$ and $\hat{\sigma}_{WR}^2$.

2. Let H_δ be the square of the absolute value of the larger of the lower and upper 90% bounds on δ derived using the t-distribution and using Satterthwaite approximation for the degrees of freedom, $H_I = \frac{\nu(\hat{\sigma}_I^2)}{\chi_\nu^2(0.05)}$, $H_T = \frac{\nu(\hat{\sigma}_{WT}^2)}{2\chi_\nu^2(0.05)}$ $H_R = \frac{-(\frac{3}{2}+2.4948)\nu(\hat{\sigma}_{WR}^2)}{\chi_\nu^2(0.95)}$ where $\chi_\nu^2(\alpha)$ is the αth-percentile point of the Chi-squared distribution with ν degrees of freedom.

3. Then

$$(\hat{\delta}^2 + \hat{\sigma}_I^2 + \frac{\hat{\sigma}_{WT}^2}{2} - (\frac{3}{2} + 2.4948)\hat{\sigma}_{WR}^2)$$

$$\mp[(H_\delta - \hat{\delta}^2)^2 + (H_I - \hat{\sigma}_I^2)^2 + (H_T - \frac{\hat{\sigma}_T^2}{2})^2 + (H_R - (-(\frac{3}{2} + 2.4948)\hat{\sigma}_{WR}^2))^2]^{\frac{1}{2}}$$

is an approximate, 90% confidence interval for the FDA's IBE metric. Appropriate modifications to this approach are made when the metric is scaled to a constant variance (see FDA Guidance, 2001).

We now find the bias and variance of the statistic $\hat{\nu}_{IBE}$ used to estimate the quantity ν_{IBE} in balanced, replicate designs with no missing data to enable application of the nonparametric bootstrap and to develop an asymptotic test for IBE.

Theorem 3.4 *Bias and Variance of the Linearised IBE Metric from FDA using Method-of-Moments*

Let

$$\hat{\nu}_{IBE} = \hat{\delta}^2 + M_I + \frac{M_T}{2} - (\frac{3}{2} + c_{FDA})M_R \tag{39}$$

be an estimate for the (38) reference-scaled metric in accordance with FDA Guidance (2001). Then, this estimate is asymptotically unbiased with

$$E[\hat{\nu}_{IBE}] = \frac{\sigma_I^2}{n} + \delta^2 + \sigma_D^2 + \sigma_{WT}^2 - (1 + c_{FDA})\sigma_{WR}^2$$

and has variance of

$$Var[\hat{\nu}_{IBE}] = \left(\frac{2\sigma_I^4}{n^2} + \frac{4\delta^2\sigma_I^2}{n} + \frac{2\sigma_I^4}{n-p}\right) + \frac{1}{n-p}(\sigma_{WT}^4 + 2\sigma_{WR}^2(\frac{3}{2} + c_{FDA})^2)$$

An unbiased reference-scaled estimator in a balanced replicate design is:

$$\hat{\nu}_{U-IBE} = \hat{\delta}^2 + (1 - \frac{1}{n})M_I + \frac{M_T}{2} - (\frac{3}{2} + c_{FDA})M_R \tag{40}$$

Then,

$$E[\hat{\nu}_{U-IBE}] = \delta^2 + \sigma_D^2 + \sigma_{WT}^2 - (1 + c_{FDA})\sigma_{WR}^2$$

and has variance of

$$Var[\hat{\nu}_{U-IBE}] = \frac{2\sigma_I^4}{n^2} + \frac{4\sigma_I^2\delta^2}{n} + \frac{2(1-\frac{1}{n})^2\sigma_I^4}{n-p} + \frac{\sigma_{WT}^4}{2(n-p)} + \frac{2\sigma_{WR}^4(\frac{3}{2}+c_{FDA})^2}{n-p}$$

Then the constant-scaled estimator:

$$\hat{\nu}_{C.IBE} = \hat{\delta}^2 + M_I + \frac{M_T}{2} - (\frac{3}{2})M_R - 0.04(c_{FDA}) \tag{41}$$

is positively biased with expectation

$$\frac{\sigma_I^2}{n} + \delta^2 + \sigma_D^2 + \sigma_{WT}^2 - \sigma_{WR}^2 - 0.04(c_{FDA})$$

and variance

$$\frac{2\sigma_I^4}{n^2} + \frac{4\sigma_I^2\delta^2}{n} + \frac{2\sigma_I^4}{n-p} + \frac{\sigma_{WT}^4}{2(n-p)} + \frac{9\sigma_{WR}^4}{2(n-p)}$$

An unbiased constant-scaled estimator is:

$$\hat{\delta}^2 + (1-\frac{1}{n})M_I + \frac{M_T}{2} - (\frac{3}{2})M_R - (0.04)c_{FDA}$$

This expression has variance:

$$\frac{2\sigma_I^4}{n^2} + \frac{4\sigma_I^2\delta^2}{n} + (1-\frac{1}{n})^2\frac{2\sigma_I^4}{n-p} + \frac{\sigma_{WT}^4}{2(n-p)} + \frac{9\sigma_{WR}^4}{2(n-p)}$$

Proof: First we note that as in Theorem 2.1, that unbiased method of moment estimators $\hat{\delta}$, M_I, M_T, and M_R for $\delta = \mu_T - \mu_R$, $\sigma_I^2 = \sigma_D^2 + \frac{\sigma_{WT}^2+\sigma_{WR}^2}{2}$, σ_{WT}^2, and σ_{WR}^2 are independent. Let (27) with corresponding assumptions hold. Then it follows directly that $\frac{\hat{\delta}^2}{\sigma_I^2/n} \sim \chi_1^{2'}(\frac{\delta^2}{\sigma_I^2/n})$ (Vonesh and Chinchilli, 1997) where $\chi_1^{2'}$ represents a non-central chi-squared distribution with one degree of freedom and non-centrality parameter $\frac{\delta^2}{\sigma_I^2/n}$ (Muirhead, p22, 1982). The expected value of $\hat{\delta}^2$ is

$$\frac{\sigma_I^2}{n}\left(1 + \frac{\delta^2}{\sigma_I^2/n}\right)$$

with variance

$$\frac{\sigma_I^4}{n^2}\left(2 + \frac{4\delta^2}{\sigma_I^2/n}\right)$$

Since $M_I \sim \frac{\sigma_I^2 \chi^2_{n-p}}{n-p}$, $M_T \sim \frac{\sigma_{WT}^2 \chi^2_{n-p}}{n-p}$, and $M_R \sim \frac{\sigma_{WR}^2 \chi^2_{n-p}}{n-p}$ are independently distributed according to the central chi-squared distribution with $n-p$ degrees of freedom, where p is the number of sequences of formulation administration. Then the expected value of M_I is σ_I^2 with variance $\frac{2\sigma_I^4}{n-p}$; the expected value of M_T is σ_{WT}^2 with variance $\frac{2\sigma_{WT}^4}{n-p}$; and the expected value of M_R is σ_{WR}^2 with variance $\frac{2\sigma_{WR}^4}{n-p}$.

Then, based on the properties of the chi-squared distribution (Muirhead, p25-27, 1982), the

$$E(\hat{\nu}_{IBE}) = \frac{\sigma_I^2}{n}(1 + \frac{\delta^2}{\sigma_I^2/n}) + \sigma_I^2 + \frac{1}{2}\sigma_{WT}^2 - (\frac{3}{2} + c_{FDA})\sigma_{WR}^2$$

$$= \frac{\sigma_I^2}{n} + \delta^2 + \sigma_I^2 + \frac{1}{2}\sigma_{WT}^2 - (\frac{3}{2} + c_{FDA})\sigma_{WR}^2$$

$$= \frac{\sigma_I^2}{n} + \delta^2 + \sigma_D^2 + \sigma_{WT}^2 - (1 + c_{FDA})\sigma_{WR}^2$$

Similarly, the proof that $E[\hat{\nu}_{U-IBE}] = \delta^2 + \sigma_D^2 + \sigma_{WT}^2 - (1 + c_{FDA})\sigma_{WR}^2$ follows from the same principles.

As each term in $\hat{\nu}_{IBE}$ are pairwise independent (see Theorem 2.1), then

$$Var(\hat{\nu}_{IBE}) = Var(\hat{\delta}^2) + Var(M_I) + \frac{1}{4}Var(M_T) + (\frac{3}{2} + c_{FDA})^2 Var(M_R)$$

from which it is found that

$$Var(\hat{\nu}_{IBE}) = \frac{2\sigma_I^4}{n^2} + \frac{4\sigma_I^2\delta^2}{n} + \frac{2\sigma_I^4}{n-p} + \frac{\sigma_{WT}^4}{2(n-p)} + (\frac{3}{2} + c_{FDA})^2 \frac{2\sigma_{WR}^4}{n-p}$$

The variance of $\hat{\nu}_{U-IBE}$ follows using the same principles. Derivations for the constant-scaled case follow similarly.□□□

Thus we now know the expected value and variance of the estimator of the linearised FDA criterion. It is possible therefore to derive asymptotic confidence intervals for the criterion and evaluate its inferential properties. This however addresses only balanced, replicate design data sets with no missing data. We now turn to REML modelling results in replicate design studies for assessing reference and constant-scaled metrics. At this point, we shall turn to asymptotic properties in REML modelling as in most data sets this will be of practical utility (i.e. when missing data is present or when imbalance exists in the data set.)

We note that as the $\hat{\nu}_{IBE}$ is a linear combination of asymptotically normal variates (Muirhead, 1982), it should be possible to use the delta-method (Serfling, 1980) and Satterthwaite's procedure (1941) or Kenward-Roger's approximations (1997) to develop a more precise confidence interval in small samples. However, as most of the data sets in IBE studies are likely to be large ($n > 28$, see Chapter 2 and FDA Guidance, 2001), the asymptotically normal interval will be explored in this thesis.

For incomplete data sets, a large sample, asymptotic hypothesis test for the linearised version of the FDA-criterion may serve as an alternative to the procedure developed by Hyslop et al. (2000). This would appear desirable given the potential for bias in Hyslop et al.'s (2000) procedure introduced by imbalance and missing data (Chapter 5). However, the degree of such bias has not been well studied, and it is not known whether the introduction of bias is of significance. We will consider how these procedures behave in practice in the retrospective analysis to be performed later in the thesis and discuss the properties of the procedures using simulation. Also to be considered is the recommended (FDA Guidance, 1997, 1999b) nonparametric-percentile application of the bootstrap (Efron and Tibshirani, 1993), and we will develop rules for when each procedure should be used and compare them.

Note that estimates arising from the use of REML models are model dependent (Chapter 2). We will consider the use of an unconstrained REML procedure (UN) and as an alternative will consider a constrained procedure (RIS).

To review, it is known (Searle, 1971) that REML estimates for the moments of interest $\delta, \sigma^2_D, \sigma^2_{WT}, \sigma^2_{WR}$ are asymptotically normally distributed with known variances. The large sample variance for $\hat{\underline{\beta}}$ and $\hat{\boldsymbol{\Sigma}}$ are $(\boldsymbol{X}'\boldsymbol{\Sigma}^{-}\boldsymbol{X})^{-}$ and $-E[\frac{\partial^2 \boldsymbol{L}^{-1}}{\partial \boldsymbol{\Sigma} \partial \boldsymbol{\Sigma}'}]$, respectively with covariance $\mathbf{0}$ where $\boldsymbol{L}$ is the *log*-likelihood in expression (30). In this situation, $\hat{\delta}^2$ will be normally distributed in the limit with expected value δ^2 and variance $4\sigma^2_{\delta}\delta^2$ where σ^2_{δ} is the usual large sample variance of δ usually estimated using the $(\boldsymbol{X}'\boldsymbol{\Sigma}^{-}\boldsymbol{X})^{-}$ matrix if $\delta \neq 0$ (Serfling, 1980). Here, $\hat{\delta}^2$ will be chi-squared distributed in the limit with expected value $\sigma^2_{\delta} + \delta^2$ and variance $2\sigma^4_{\delta} + 4\sigma^2_{\delta}\delta^2$ where σ^2_{δ} is the usual large sample variance of δ usually estimated using the $(\boldsymbol{X}'\boldsymbol{\Sigma}^{-}\boldsymbol{X})^{-}$ matrix if $\delta = 0$. In bioequivalence testing, it is most of interest to consider the first situation, and we will develop the properties of the estimators under this condition. We will later use simulation to ensure that the procedure we develop operates adequately under the second condition.

We now turn to the variance estimates. These are normally distributed in the limit with variance-covariance matrix appropriate to the structure of the model. For a REML UN model, where we estimate $\begin{pmatrix} \hat{\sigma}^2_{BT} \\ \hat{\omega}_{RT} \\ \hat{\sigma}^2_{BR} \\ \hat{\sigma}^2_{WT} \\ \hat{\sigma}^2_{WR} \end{pmatrix}$, these are normally distributed with expected value

$\begin{pmatrix} \sigma^2_{BT} \\ \omega_{RT} \\ \sigma^2_{BR} \\ \sigma^2_{WT} \\ \sigma^2_{WR} \end{pmatrix}$ and symmetric variance-covariance of

$-E[\frac{\partial^2 \boldsymbol{L}^{-1}}{\partial \boldsymbol{\Sigma} \partial \boldsymbol{\Sigma}'}]$, the terms of which we shall denote as

$\begin{pmatrix} l_{BT} & l_{BTx\omega} & l_{BTxBR} & l_{BTxWT} & l_{BTxWR} \\ l_{BTx\omega} & l_{\omega} & l_{BRx\omega} & l_{\omega xWT} & l_{\omega xWR} \\ l_{BTxBR} & l_{BRx\omega} & l_{BR} & l_{BRxWT} & l_{BRxWR} \\ l_{BTxWT} & l_{\omega xWT} & l_{BRxWT} & l_{WT} & l_{WTxWR} \\ l_{BTxWR} & l_{\omega xWR} & l_{BRxWR} & l_{WTxWR} & l_{WR} \end{pmatrix}$

Similarly, for the constrained REML model RIS, we find that where we estimate $\begin{pmatrix} \hat{\sigma}^2_D/2 \\ \hat{\sigma}^2_{WT} \\ \hat{\sigma}^2_{WR} \end{pmatrix}$, these are normally distributed with expected value

$\begin{pmatrix} \sigma^2_D/2 \\ \sigma^2_{WT} \\ \sigma^2_{WR} \end{pmatrix}$ and symmetric variance-covariance of

$-E[\frac{\partial^2 \boldsymbol{L}^{-1}}{\partial \boldsymbol{\Sigma} \partial \boldsymbol{\Sigma}'}]$, the terms of which we shall denote as

$\begin{pmatrix} l_D & l_{DxWT} & l_{DxWR} \\ l_{DxWT} & l_{WT} & l_{WTxWR} \\ l_{DxWR} & l_{WTxWR} & l_{WR} \end{pmatrix}$

From these definitions, it is easy to derive the expected values and variances of the relevant estimators of the FDA metric. The delta method may then be applied to construct asymptotically normal confidence intervals for the metric.

Theorem 3.5 *Asymptotic Bias and Variance of the Linearised IBE Metric from FDA using REML Estimation*

Let

$$\hat{\nu}_{IBE} = \hat{\delta}^2 + \hat{\sigma}^2_{BT} + \hat{\sigma}^2_{BR} - 2\hat{\omega}_{RT} + \hat{\sigma}^2_{WT} - (1 + c_{FDA})\hat{\sigma}^2_{WR} \tag{42}$$

be an estimate for the (38) reference-scaled metric in accordance with FDA Guidance (2001) using a REML UN model. Then, this estimate is asymptotically normally distributed, unbiased with

$$E[\hat{\nu}_{IBE}] = \delta^2 + \sigma^2_D + \sigma^2_{WT} - (1 + c_{FDA})\sigma^2_{WR}$$

and variance of

$$Var[\hat{\nu}_{IBE}] = 4\sigma^2_{\delta}\delta^2 + l_{BT} + l_{BR} + 4l_{\omega} + l_{WT} + (1 + c_{FDA})^2(l_{WR}) +$$

$$2l_{BTxBR} - 4l_{BTx\omega} + 2l_{BTxWT} - 2(1 + c_{FDA})l_{BTxWR} - 4l_{BRx\omega} + 2l_{BRxWT} -$$

$$2(1+c_{FDA})l_{BRxWR} - 4l_{\omega xWT} + 4(1+c_{FDA})l_{\omega xWR} - 2(1+c_{FDA})l_{WTxWR}$$

Let

$$\hat{\nu}_{C.IBE} = \hat{\delta}^2 + \hat{\sigma}^2_{BT} + \hat{\sigma}^2_{BR} - 2\hat{\omega}_{RT} + \hat{\sigma}^2_{WT} - \hat{\sigma}^2_{WR} - 0.04(c_{FDA}) \tag{43}$$

be an estimate for the constant-scaled metric in accordance with FDA Guidance (2001) using a REML UN model. Then, this estimate is asymptotically normally distributed, unbiased with

$$E[\hat{\nu}_{C.IBE}] = \delta^2 + \sigma^2_D + \sigma^2_{WT} - \sigma^2_{WR} - 0.04(c_{FDA})$$

and has variance of

$$Var[\hat{\nu}_{C.IBE}] = 4\sigma^2_\delta\delta^2 + l_{BT} + l_{BR} +$$

$$4l_\omega + l_{WT} + l_{WR} + 2l_{BTxBR} - 4l_{BTx\omega} + 2l_{BTxWT} - 2l_{BTxWR} - 4l_{BRx\omega} +$$

$$2l_{BRxWT} - 2l_{BRxWR} - 4l_{\omega xWT} + 4l_{\omega xWR} - 2l_{WTxWR}$$

Similarly, let

$$\hat{\nu}_{IBE} = \hat{\delta}^2 + 2(\hat{\sigma}^2_D/2) + \hat{\sigma}^2_{WT} - (1+c_{FDA})\hat{\sigma}^2_{WR} \tag{44}$$

be an estimate for the (38) reference-scaled metric in accordance with FDA Guidance (2001) and using a REML RIS model. Then, this estimate is asymptotically normally distributed, unbiased with

$$E[\hat{\nu}_{IBE}] = \delta^2 + \sigma^2_D + \sigma^2_{WT} - (1+c_{FDA})\sigma^2_{WR}$$

and has variance of

$$Var[\hat{\nu}_{IBE}] = 4\sigma^2_\delta\delta^2 + 4l_D + l_{WT} + (1+c_{FDA})^2 l_{WR} +$$

$$4l_{DxWT} - 4(1+c_{FDA})l_{DxWR} - 2(1+c_{FDA})l_{WTxWR}$$

Let

$$\hat{\nu}_{C.IBE} = \hat{\delta}^2 + 2(\hat{\sigma}^2_D/2) + \hat{\sigma}^2_{WT} - \hat{\sigma}^2_{WR} - 0.04(c_{FDA}) \tag{45}$$

be an estimate for the constant-scaled metric in accordance with FDA Guidance (2001) using a

REML RIS model. Then, this estimate is asymptotically normally distributed, unbiased with

$$E[\hat{\nu}_{C.IBE}] = \delta^2 + \sigma^2_D + \sigma^2_{WT} - \sigma^2_{WR} - 0.04(c_{FDA})$$

and has variance of

$$Var[\hat{\nu}_{C.IBE}] = 4\sigma^2_\delta\delta^2 + 4l_D + l_{WT} + l_{WR} + 4l_{DxWT} - 4l_{DxWR} - 2l_{WTxWR}$$

Proof: Here we apply the findings of Theorem 3.3.A of Serfling (1980) using the properties described previously of the estimates making up $\hat{\nu}_{IBE} = g(\hat{\delta}, \hat{\sigma}^2_{BT}, \hat{\sigma}^2_{BR}, \hat{\omega}, \hat{\sigma}^2_{WT}, \hat{\sigma}^2_{WR})$. The function g is obviously differentiable such that $\frac{\partial g}{\partial \hat{\delta}}\Big|_{g=\delta} = 2\delta$, $\frac{\partial g}{\partial \hat{\sigma}^2_{BT}}\Big|_{g=\sigma^2_{BT}} = 1$, $\frac{\partial g}{\partial \hat{\sigma}^2_{BR}}\Big|_{g=\sigma^2_{BR}} = 1$, $\frac{\partial g}{\partial \hat{\omega}}\Big|_{g=\omega} = -2$, $\frac{\partial g}{\partial \hat{\sigma}^2_{WT}}\Big|_{g=\sigma^2_{WT}} = 1$, and $\frac{\partial g}{\partial \hat{\sigma}^2_{WR}}\Big|_{g=\sigma^2_{WR}} = -(1 + c_{FDA})$.

Then by application of Theorem 3.3.A (Serfling, 1980), it is found that $g(\hat{\delta}, \hat{\sigma}^2_{BT}, \hat{\sigma}^2_{BR}, \hat{\omega}, \hat{\sigma}^2_{WT}, \hat{\sigma}^2_{WR})$ is asymptotically normally distributed with expected value $g(\delta, \sigma^2_{BT}, \sigma^2_{BR}, \omega, \sigma^2_{WT}, \sigma^2_{WR})$ and variance $\underline{D}\boldsymbol{\Sigma}_l\underline{D}'$ where $\underline{D} = (2\delta, 1, 1, -2, 1, -(1 + c_{FDA}))$ and where $\boldsymbol{\Sigma}_l$ is the Unstructured REML asymptotic variance-covariance matrix above augmented with the σ^2_δ associated in the first row, first column, such that

$$\boldsymbol{\Sigma}_l = \begin{pmatrix} \sigma^2_\delta & 0 & 0 & 0 & 0 & 0 \\ 0 & l_{BT} & l_{BTx\omega} & l_{BTxBR} & l_{BTxWT} & l_{BTxWR} \\ 0 & l_{BTx\omega} & l_\omega & l_{BRx\omega} & l_{\omega xWT} & l_{\omega xWR} \\ 0 & l_{BTxBR} & l_{BRx\omega} & l_{BR} & l_{BRxWT} & l_{BRxWR} \\ 0 & l_{BTxWT} & l_{\omega xWT} & l_{BRxWT} & l_{WT} & l_{WTxWR} \\ 0 & l_{BTxWR} & l_{\omega xWR} & l_{BRxWR} & l_{WTxWR} & l_{WR} \end{pmatrix}$$. The proof then follows by matrix multiplication. The proofs for the other situations described above follow in the same manner and are not reproduced here.□□□

These findings allow for the derivation of approximate confidence intervals for the linearised FDA criterion in a straightforward manner as we know (Serfling, 1980) that

$$\frac{g(\hat{\delta}, \hat{\sigma}^2_{BT}, \hat{\sigma}^2_{BR}, \hat{\omega}, \hat{\sigma}^2_{WT}, \hat{\sigma}^2_{WR}) - g(\delta, \sigma^2_{BT}, \sigma^2_{BR}, \omega, \sigma^2_{WT}, \sigma^2_{WR})}{\sqrt{\underline{D}\hat{\boldsymbol{\Sigma}}_l\underline{D}'}} \underrightarrow{d} N(0,1)$$

where $\underrightarrow{d}$ denotes convergence in distribution and $\hat{\boldsymbol{\Sigma}}_l$ is the estimated asymptotic variance-covariance matrix (observed inverse Fisher information), a consistent estimate for $\boldsymbol{\Sigma}_l$.

We now have straightforward means of assessing the average and individual bioequivalence between formulations using a variety of techniques. Bias in the estimates arising from Method-of-Moments estimation will be considered, and the use of REML versus Method-of-Moments

estimation for IBE assessment under recommendations of relevant regulatory guidance and using the constrained and unconstrained REML asymptotic procedures described above will be studied using the 51 replicate design data sets described previously (Section 2.7.1).

Unfortunately, as we have seen in the previous Chapter, complete data sets are a rarity in bioequivalence studies, and Method-of-moments estimation is severely limited in such situations (as will be established in Chapter 5). We will investigate whether REML asymptotic procedures yield reasonable conclusions in this context in small samples using simulation and discuss findings from modest simulations carried out to examine the sensitivity of $\hat{\sigma}_D^2$ in Chapter 5.

3.4 Retrospective Analysis

We now turn to discussion of retrospective analysis of the replicate design data sets for the assessment of individual bioequivalence.

Three inferential methods were explored:

1. The approximation-method (Hyslop et al., 2000) based on the Cornish-Fisher expansion was applied to the data sets in accordance with FDA Guidance (2001) using method-of-moments estimates to derive the upper bound of a 90% confidence interval for the linearised FDA population bioequivalence metric.

2. The non-parametric percentile method (Efron and Tibshirani, 1993) was used with 2000 bootstraps, maintaining the observed numbers of subjects per sequence, and using an unrestricted (UN) REML model to derive the upper bound of a 90% confidence interval for the linearised FDA population bioequivalence metric. The unrestricted model was selected instead of a restricted REML model (recommended in FDA Guidance, 1997) in order to provide consistency relative to the method-of-moments procedure used in the Hyslop et al (2000) method described above. Method-of-moments and unrestricted REML estimates should be equal when the data set is strongly balanced (Vonesh and Chinchilli, 1997) and has no missing data. Also this model was selected in order to provide consistency with the asymptotic procedure developed in previously in this Chapter.

3. The asymptotic procedure developed previously in this Chapter was applied to each data set using an unrestricted REML model and a restricted REML model.

We first describe the results of each analysis and then compare and contrast between them.

This sub-section ends with discussion of significant findings to be explored through the use of simulation and conclusions. Results of the Hyslop et al. (2000) analysis may be found in Tables 54-55. Results of analysis using the REML asymptotic test developed in this thesis may be found in Tables 59 and 61. Nonparametric bootstrap findings may be found in Tables 57-58.

Many data sets failed to demonstrate IBE. The procedure of Hyslop et al. (2000) found that nine AUC data sets (G, I1, I2, J, Q2, R, W4, ZB, and ZE1) and 14 Cmax data sets (D, E, F, G, I1, I2, M, N1, N2, P, Q2, W3, X, and ZB) failed to demonstrate IBE. The bootstrap procedure found that 11 AUC data sets (G, I1, I2, Q2, R, T, V, W4, W6, ZB, and ZE1) and 17 Cmax data sets (D, E, F, G, I1, I2, M, N1, O1, O2, Q2, T, W3, X, Y, ZB, and ZE1) failed to demonstrate IBE while the unrestricted asymptotic procedure finds that 11 AUC data sets (G, I1, I2, J, O2, Q2, R, W4, W6, ZB, and ZE1) and 16 Cmax data sets (D, E, F, G, I1, I2, J, M, N1, O2, P, Q2, W3, X, ZB, and ZE1) failed to demonstrate IBE.

For AUC data sets R and W6, it was evident that accounting for the non-null the covariance between variance estimates using a bootstrap or asymptotic procedure, differing inference can result relative to a procedure which assumes it to be null (Hyslop et al., 2000). For the remainder of the AUC data sets where a discrepancy is observed (J, O2, T, and V), no reason was readily evident, and we will use simulation to explore the error rates subsequently in this thesis. Constraint on the variance estimates may play a role in this setting for data set J as when the REML model is constrained to the usual parameter space, IBE was demonstrated.

For Cmax, discrepancies were observed between procedures for data sets J, N2, P, T, Y, and ZE1. For data sets N2 and ZE1, it was evident that non-null correlation again contributes to passage or failure for asymptotic and bootstrap procedures while the Hyslop et al. (2000) procedure fails to account for these factors. Data set P passes but was very near the cut-off of zero when using the bootstrap; it is likely that is more bootstraps were added, this data set would fail to demonstrate IBE. Thus in three data sets, no reason for the discrepancy was readily evident and may be attributable to random error.

Constraint of the parameter space did impact the results of the asymptotic procedure for Cmax. Data sets E, F, J, and N2 failed under the unrestricted asymptotic procedure however, these data sets demonstrated IBE under the constrained REML model.

Lastly, we considered the estimated bias in the individual bioequivalence metric, equal as

previously established in this Chapter as σ_I^2/n. We neglect that this bias is exact for only strongly balanced, complete data sets for the purposes of this exercise. The mean (STD) bias for AUC and Cmax were estimated as 0.002 (0.001) and 0.004 (0.003), respectively. Expressed as an absolute percentage of the estimate for the IBE metric, this mean (STD) was 1.50% (4.61%) and 2.19% (3.83%), respectively. Accounting for this positive bias was not found to impact inference relative to the Hyslop upper bound in any data set, as would be expected given the asymptotic unbiasedness of the metric previously established in this Chapter relative to the sample sizes employed in these studies.

3.5 Discussion and Findings

We begin discussion with consideration of the goalposts chosen by FDA for acceptance of IBE. The goalpost allows for large changes in average exposure (which would result in ABE) failure due to decreased within-subject variability on the test formulation relative to reference or by scaling to large within-subject reference variation. Based on the findings of retrospective analysis, we believe this goalpost may be too liberal and has the potential to endanger public health. We will study this potential issue using simulation in Chapter 5.

The procedure developed by Hyslop et al. (2000) is flawed in a theoretical sense due to its assumption of independence among the estimates making up the metric. This assumption is violated when missing data or imbalance are present in the data. However, when the asymptotic test and nonparametric bootstrap were applied (which do account for these non-null correlations), only slight differences in inference were observed. We will study this question further in Chapter 5.

Similar to the PBE findings (as will be seen in Chapter 4), several data sets exhibit the trade-offs described previously (Zariffa et al., 2000) allowing for demonstration of IBE and presumably market access. These trade-offs appear undesirable in practice in order to ensure equal therapeutic benefit in the marketplace.

While both REML and method-of-moment procedures were shown to provide asymptotically unbiased estimates in replicate design studies (under certain conditions), slight discrepancies in IBE inference were observed when using the Hyslop et al. (2000), Asymptotic, and nonparametric bootstrap methods. Some seemingly random discrepancies are also observed, as is not

unexpected in a data set this large ($n = 51$). We will explore the relative Type I and II error rates using simulation in Chapter 5 following an exploration using simulation of the properties of σ_D^2, of particular concern in IBE assessment and the potential for bias in small sample estimates from the various models.

In conclusion, we find that for the majority of data sets, the Hyslop et al. (2000), bootstrap, and asymptotic procedures yield the same conclusions with regard to the data sets analysed. Bias in small sample estimates (not accounting for bas introduced by missing data) is minor in these data sets and was not observed to impact inference. Minor discrepancies are observed which may be due to the practice of accounting for correlation among the estimated moments, to constraints placed on the variance estimates themselves (i.e. to be non-zero), to bias introduced by missing data, and due to differences in Type I and II error rates of these testing procedures studied.

4 Small and Large Sample Properties, Estimation, and Inference for Population Bioequivalence

The findings of this chapter were presented at the annual American Society of Clinical Pharmacology and Therapeutics meeting (Patterson et al., 2000a), at the American Association of Pharmaceutical Scientists joint workshop with the USA Food and Drug Administration (Zariffa and Patterson, 2000), at the Societe de Statistique Francaise (Patterson et al., 2001f), at the American Statistical Society Joint Statistical meetings (Patterson and Jones, 2002e), and at the International Society of Clinical Biostatistics meeting (Patterson and Jones, 2002f). Aspects of the findings were published in the *Journal of Clinical Pharmacology* (Zariffa and Patterson, 2001), in *Pharmaceutical Statistics* (Patterson and Jones, 2002a; 2002g), in the *Proceedings of the Joint Statistical Meetings* (Patterson and Jones, 2002h), and in a series of a GlaxoSmithKline technical reports (Patterson et al., 2001e; Patterson and Jones, 2002b-c).

4.1 Introduction and Goals of Chapter

In this Chapter, key ideas in population bioequivalence will be quickly reviewed, and we then turn to detailed assessment of the properties of population bioequivalence.

To review, average bioequivalence (ABE; FDA Guidance, 1992) has traditionally been used as the standard for market access with regulatory limits of twenty percent. This approach and the models used to test for ABE are described in Chapter 2. In practical terms, a ninety percent confidence interval is constructed on the $\hat{\mu}_T - \hat{\mu}_R$. If the confidence intervals for both AUC and Cmax fall within the range $-\ln 1.25$ to $\ln 1.25$, then average bioequivalence is demonstrated. More commonly, these differences and confidence intervals are exponentiated and assessed relative to the interval 0.80 to 1.25.

Following the original proposal in 1997 (cf. FDA Guidance), in August 1999, the FDA re-proposed new guidelines for the assessment of bioequivalence: population bioequivalence (PBE) and individual bioequivalence (IBE) (FDA Guidances, 1999a and 1999b) and finalised procedures in 2000-2001 (FDA Guidance, 2000b, 2001) based on ideas developed by Anderson (1993) and Anderson and Hauck (1983, 1990). In the case of pre-market approval, one can formulate the bioequivalence question as "Can a patient begin their therapy with either formulation (commercial or clinical trial) and be assured same results in terms of safety and efficacy?" This has been called the concept of prescribability (Anderson and Hauck, 1990) and is linked to PBE. PBE has not been well studied (see Chapter 1) as most attention has focussed on IBE (the subject of Chapter 2), and we will develop methods appropriate to its study in cross-over designs in this thesis.

Population bioequivalence is assessed using the following aggregate statistic (FDA Guidance, 1997).

$$\frac{(\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2}{max(0.04, \sigma_R^2)} \tag{46}$$

where $\sigma_T^2 = \sigma_{WT}^2 + \sigma_{BT}^2$ and $\sigma_R^2 = \sigma_{WR}^2 + \sigma_{BR}^2$. Note that this aggregate statistic can be constructed using a mixed model from a two period cross-over design (with appropriate modification to model (22)) and does not require the use of a replicate design.

In the case of post marketing changes, the bioequivalence question becomes: "Can I safely and effectively switch my patient from their current formulation to another?" This has been called the concept of switchability (Anderson and Hauck, 1990) and is linked to IBE. The crite-

ria used to assess IBE under the proposed FDA draft guidance (1997-1999) and finalised procedures in 2000-2001 (FDA Guidance, 2000b, 2001) aggregates the difference between population means and variances and accounts for subject predictability from one formulation to the other (subject-by-formulation interaction, Ekbohm and Melander, 1989). In addition, the individual bioequivalence metric allows for scaling of the regulatory limits based on the within-subject variability of the reference product.

Previous findings on PBE are described in Zariffa et al. (2000a) and in the previous Chapter 3 and will not be reviewed here. PBE has been neglected in favor of research in IBE in recent years and little attention (Chapter 1; see also Barrett et al., 2000) has focussed on the use of such a metric. This situation will be rectified in this thesis.

4.2 Estimation Methods and Inferential Procedures for Population Bioequivalence

We now begin discussion of population bioequivalence, neglected for the most part in the statistical and biomedical literature in favor of the study of individual bioequivalence.

4.2.1 Estimation Procedures

In the assessment of population bioequivalence metrics, it will be of interest to compare various properties of first and second order moments (Rao, 1973). Such methods are well described in the statistical literature for parallel group designs (the subject of Chapter 6), and such will not be discussed here. Instead we will dwell on estimation methods used in cross-over studies for the assessment of population bioequivalence.

For the two-period cross-over, methods are utilised which are consistent with the methods described in Jones and Kenward (1989), Senn (1993), Vonesh and Chinchilli (1997), and Senn (2002) making use of the bi-variate normal (in large samples or when using asymptotic theory) or the bi-variate t-distribution in small samples.

The following mixed model for log_e-transformed observations can be applied in replicate designs. Let $y_{ijk(t)}$ be the response for the k-th subject ($k = 1, 2, \ldots, n_i$) in sequence i ($i = 1, 2, \ldots, s$) and period j ($j = 1, 2$) in the cross-over trial administered formulation t ($t = T, R$)

and

$$y_{ijk(t)} = \lambda_i + \pi_j + \mu_t + \varepsilon_{ijk(t)} \tag{47}$$

where λ_i and π_j are nuisance effects (sequence and period effects),

$\varepsilon_{ijk(t)}$ are normally distributed with mean zero

$Var(\varepsilon_{tjk}) = \sigma_t^2 = \sigma_{Wt}^2 + \sigma_{Bt}^2$, the sum of between- and within-subject variance,

$Cov(\varepsilon_{ijk(T)}, \varepsilon_{ij'k(t')}) = \omega_{RT}$, for $j \neq j'$. Subjects are assumed to be independent. Under this model $Var(y_{ijk(t)} - y_{ij'k(t')}) = \sigma_T^2 + \sigma_R^2 - 2\omega_{RT} = \sigma_{WT}^2 + \sigma_{BT}^2 + \sigma_{WR}^2 + \sigma_{BR}^2 - 2\omega_{RT}$, the subject-by-formulation interaction variance plus the sum of within-subject variances (Senn, 1993; Vonesh and Chinchilli, 1997; Senn, 2002). If the Huyhn-Feldt condition is applied, this is the model of Jones and Kenward (1989).

Standard analysis of variance (Rao, 1973; Searle, 1971) may be applied if the Huyhn-Feldt condition (1970) holds to generated unbiased estimators for the between and within-subject variances and best-linear unbiased estimates for the means of each formulation in balanced two-period cross-overs; furthermore, restricted maximum likelihood models may be applied to derive asymptotically unbiased estimates for the moments of interest (Jones and Kenward, 1989; Milliken and Johnson, 1992). Method of moments estimation is described in Vonesh and Chinchilli (1997) when this condition is not applied, and will not be reproduced here. Here, an approach is developed using PROC MIXED ($SAS^{©}$) to provide an example of the principles in unbalanced models.

In matrix notation, the above model can be expressed as

$$\underline{y} = \boldsymbol{X}\underline{\beta} + \boldsymbol{Z}\underline{u} \tag{48}$$

where $\underline{u}$ is a multivariate normal density (MVN) with expectation $\underline{0}$ and variance-covariance matrix $\boldsymbol{\Omega}$ ($\underline{u} \sim MVN(\underline{0}, \boldsymbol{\Omega})$ and $'$ denotes the transpose of a matrix. $\boldsymbol{\Omega}$ will be composed of a $4n \times 4n$ matrix of variances and covariances $\boldsymbol{Z\Omega Z'} = \boldsymbol{\Omega}$. Let $\boldsymbol{\Omega}$ be defined from the following set of variance-covariance components $(\sigma_R^2, \omega_{RT}, \sigma_T^2)$ corresponding to the method-of-moments approach (Chinchilli and Esinhart, 1996) where ω_{RT} represents the covariance between test and reference observations under model (47). $\boldsymbol{Z}$ is a matrix of composed from the set $(0, 1)$ to assemble the covariance matrix of observations in a manner appropriate to sequence i for each

subject j (subjects are assumed to be independent. PROC MIXED ($SAS^{©}$) code may then be written to derive asymptotically unbiased estimates for the moments of interest.

```
proc mixed method=reml;
class sequence subject period regimen;
model lnauc=sequence period regimen;
repeated regimen/subject=subject type=UN;
run;
```

This REML model will be applied later in this Chapter to provide an example of the application of the estimation technique in a two-period cross-over. Estimation methods in a replicate design are described the preceding sections.

4.2.2 Properties of the Estimated Metrics for PBE Assessment

We now turn to consideration of the population bioequivalence metric from FDA (FDA Guidance, 1997, 1999a-b, 2000b).

To review, summary measurements such as AUC from a two-by-two cross-over trial may be modeled using a random-intercept mixed modelling procedure accounting for each subject as their own control (Jones and Kenward, Ch 7, 1989; Milliken and Johnson, Ch 32, 1992). In bioequivalence studies, the following model for observations is commonly accepted for a randomized, two period, cross-over trial in normal healthy volunteers, under the assumption that carryover effects are negligible (or similar between formulations). Let y_{ijk} be the $\log_e$-transformed j-th period's observation ($j = 1, 2$) for the k-th subject ($k = 1, 2, \ldots, n_i$) in the i-th sequence group ($i = 1, 2$). Then

$$y_{ijk} = \lambda_i + (\mu + \nu_{k(i)}) + \pi_j + \tau_{d[i,j]} + \varepsilon_{ijk}$$

where μ is the grand mean,

λ_i, π_j, and $\tau_{d[i,j]}$ are fixed effects for sequence, period, and formulation,

$\nu_{k(i)}$ and ε_{ijk} are random effects which are independent with mean zero, $\mathrm{Var}(\nu_{k(i)})=\sigma_B^2$, the between-subject variance, and

$\text{Var}(\varepsilon_{ijk})=\sigma^2_W$, the within-subject variance. This equation is usually expressed in practice as:

$$y_{ijk} = \lambda_i + \nu_{k(i)} + \pi_j + \mu_d + \varepsilon_{ijk}$$

where $d = (T, R)$. Analysis of data under $\log_e$-transformation is described in Box and Cox (1964). See also Chapter 1.

For balanced designs (n_1=n_2) with no missing data, period effects are orthogonal (non-aliased) with formulation effects. Note that homogeneity of test and reference product between-subject variance is assumed in such a random-intercept model, as is the homogeneity of within-subject variance for test and reference products. Observations between-subject are held to be independent (with covariance zero), and the variance-covariance structure for observations within a subject under these assumptions is compound symmetric. Under this model, the $Cov(y_{ijk}, y_{ij'k}) = E(\nu_{k(i)} + \varepsilon_{ijk})(\nu_{k(i)} + \varepsilon_{ij'k}) = E(\nu^2_{k(i)}) = Var(\nu_{k(i)}) = \sigma^2_B$ such that:

$$\rho = Corr(y_{ijk}, y_{ij'k}) = \frac{\sigma^2_B}{\sigma^2_B + \sigma^2_W}$$

for $j \neq j'$, and

$$\sigma^2 = Var(y_{ijk}) = \sigma^2_B + \sigma^2_W$$

Comparisons between the estimated means $\hat{\mu}_T$-$\hat{\mu}_R$ are thus (Jones and Kenward, 1989) normally-distributed with mean μ_T-μ_R and variance of $((\sigma^2_B+\sigma^2_W)+(\sigma^2_B+\sigma^2_W)-2(\rho)(\sqrt{\sigma^2_B+\sigma^2_W})(\sqrt{\sigma^2_B+\sigma^2_W}))/n = 2(\sigma^2_W)/n$ in balanced two-period cross-over designs with no missing data and n subjects. Estimates of variance may be derived using method-of-moments or REML estimation, and these estimates are unbiased (or asymptotically unbiased in the REML case) for the true variances. Tests of fixed effects are exact under the Huyhn-Feldt condition, and may be constructed in the usual fashion (see Chapter 1) for the assessment of average bioequivalence (FDA Guidance 1992, 2001).

Approaches to *scaled* average bioequivalence have been described previously (Endrenyi, 1994;

Dragalin and Fedorov, 1999a). In situations with missing data, REML modelling may be used to derive asymptotic inference, or the bootstrap may be used to assess inference. Regulatory bodies have not been readily receptive of the method in practice (FDA Guidance, 2001) given the uncertainty in precision in variance components and as this criteria still ignores differences in variability between formulations and subject-by-formulation interaction. Discussion of this approach thus will be neglected in favor of methods potentially of relevance to implementation in the pharmaceutical industry.

Considering an alternative expression of the data in a balanced two-period cross-over of n subjects with sequences i (i =TR,RT), let $\begin{pmatrix} \bar{y}_{Ri} \\ \bar{y}_{Ti} \end{pmatrix} \sim MVN_2 \begin{pmatrix} \mu_T + p_{i1} \\ \mu_R + p_{i2} \end{pmatrix}, \begin{pmatrix} \sigma_R^2 & \omega_{RT} \\ \omega_{RT} & \sigma_T^2 \end{pmatrix}$ where p_{i1} and p_{i2} denote the effects of period appropriate to sequence i for each subject. Similarly, mean and variance estimates may be constructed for replicate designs. In unbalanced data sets, REML models may be used to derive asymptotically unbiased estimates.

Pitman (1939) and Morgan (1939) approaches for equality of variances arising from paired data may be found in Chow and Liu (Chapter 7, 2000) and will not be reproduced in this thesis.

In two-period or four-period replicate designs, the Huyhn-Feldt condition (1970) for variances across formulations may not be applicable. Vonesh and Chinchilli (1997) present methods for the estimation of total, between, and within-subject variances in replicate and cross-over designs. Here we note only that careful consideration of the covariance between observation must be implemented when analysing such data so as to prevent the classic errors associated with the analysis of cross-over data (Senn, 1993; Senn, 2002). Note that estimates σ_T^2 and σ_R^2 in a two-period cross-over and estimates $\sigma_t^2 = \sigma_{Bt}^2 + \frac{\sigma_{Wt}^2}{2}$ (for $t = T, R$ representing Test and Reference formulations) in replicate designs should be correlated in a cross-over design (Jones and Kenward, 1989).

This finding impacts the use of the proposed FDA procedure for PBE (FDA Guidance, 2001) as this intrinsic correlation is not taken into account when forming the approximate confidence interval (Hyslop et al., 2000) for the linearised criterion

$$\nu_{PBE} = \delta^2 + \sigma_T^2 - (1 + c_{FDA})\sigma_R^2 \tag{49}$$

or

$$\nu_{C.PBE} = \delta^2 + \sigma_T^2 - \sigma_R^2 - c_{FDA}(0.04) \tag{50}$$

It is easy to show that an unbiased method-of-moments (or an asymptotically unbiased) estimate $\hat{\nu}_{PBE}$ for ν_{PBE} may easily be derived using unbiased estimates from a parallel group, four-period replicate cross-over, or two-period cross-over design, as we will study further below.

As in Chapter 3, the FDA (2001) guidance specifies that both the constant and reference scaled metrics should be constructed independent of the level of reference product variation and the approach to estimation developed in this thesis were studied under this approach. Hence the expectations and variances are derived independent of the level of estimated total reference variation for PBE. Impact on PBE statistics would be expected to be negligible in this context as total variability is in general well above the 0.04 level in PK studies.

We first turn to consideration of the goalpost in PBE assessment. The goalpost for population bioequivalence (see Chapter 1) assessment assumes the variance for the reference formulation is 0.04. The difference $\mu_T - \mu_R$ is allowed to take on a value of up to ln 1.25 and a variance allowance of 0.5 in the numerator under the procedure proposed by the FDA (cf. FDA Guidance, 1999). Thus the regulatory 'cut-off' is constrained to a level of 1.74. If the upper 95% bound on the FDA metric falls below this value of 1.74, population bioequivalence is demonstrated for the endpoint under study. Scaled to reference variation (again, assumed to be 0.04, under the FDA Guidance 1997), the goalpost accounting for the means amounts to a value of $1.24{=}(\ln(1.25))^2/0.04$. The remaining allowance, known as the 'variance allowance' and is equal to 0.5 (allowing for a difference in variances $((\sigma_T^2 - \sigma_R^2)/\sigma_R^2){=}(0.02/0.04)$, when scaled to reference product variation.)

One questions whether this is an appropriate choice of goalpost when looking at a PBE study. We illustrate this point by separating the metric into components $x = \frac{(\mu_T - \mu_R)^2}{\sigma_R^2}$ and $y = \frac{\sigma_T^2}{\sigma_R^2}$. Note that (46) is equal to $x + y - 1$.

In Figure 20, the response-surface for combinations of x and y yielding (46) is plotted both as a surface and, in the second part of the figure, as a projection onto the plane of possible responses as a function of x and y. Here we see the potential for trade-offs in (46). For example, in most cases, the value on the x-axis will be close to zero (see Retrospective analysis later in this Chapter). In such cases, we observed that the value on the y-axis can vary quite dramatically, yet still fall below the 1.74 acceptance value. However, at the present time 1.74 has been selected as the relevant value, and we will utilise it in further research.

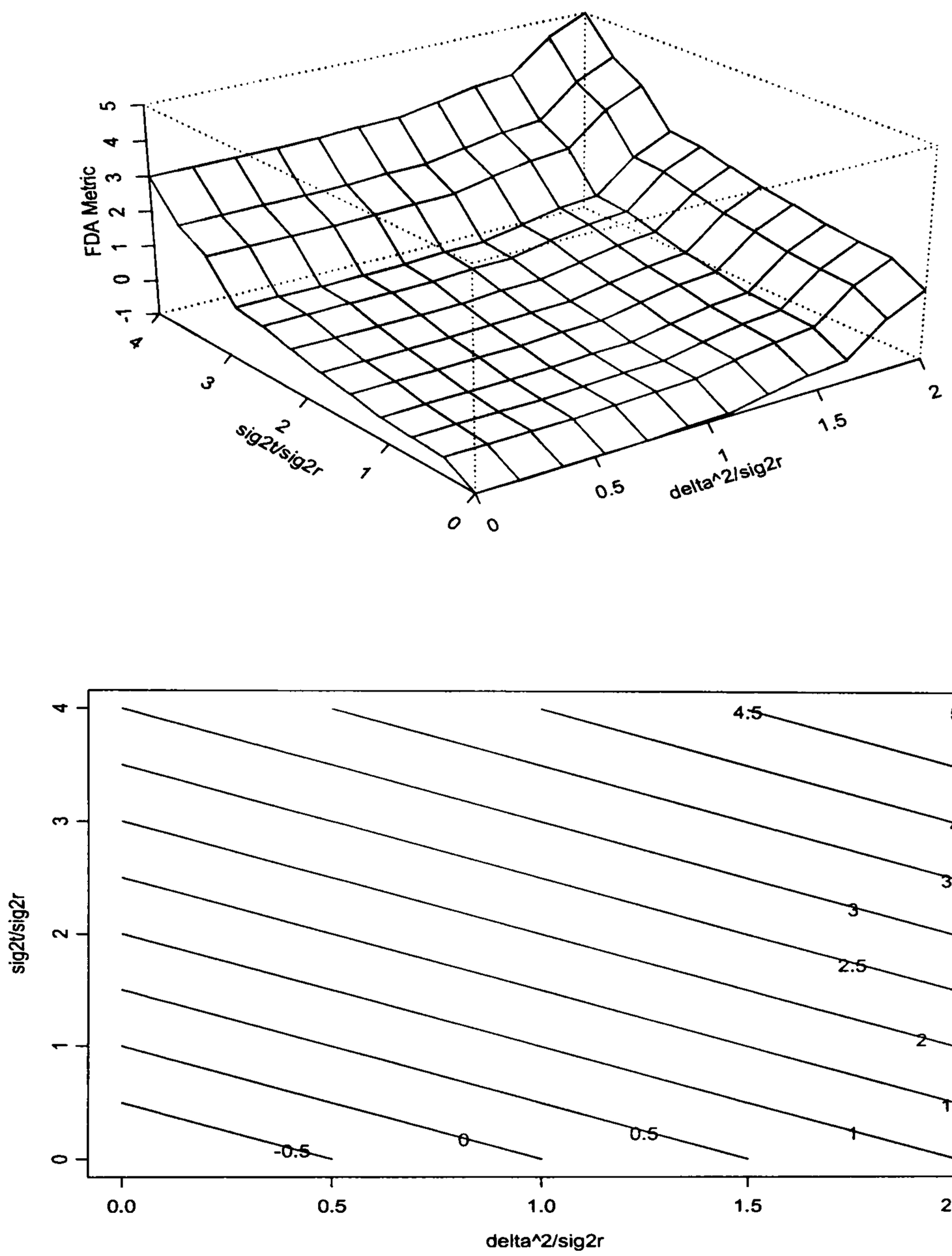


Figure 20: Response-Surface and Projected Values of (46)$= x + y - 1$ relative to $x = \frac{(\mu_T - \mu_R)^2}{\sigma_R^2}$ and $y = \frac{\sigma_T^2}{\sigma_R^2}$

In the findings below, we will show that method-of-moments and REML estimation result in asymptotically normally distributed estimates with estimable variance. These findings are important as they allow the use of the nonparametric bootstrap and asymptotic tests in the assessment of PBE inference and support the use of the FDA Guidance (2001) procedure in the assessment of inference.

In a two-period cross-over, independent unbiased method-of-moments estimators $\hat{\delta}$, $\hat{\sigma}_T^2$, and $\hat{\sigma}_R^2$ are derived for $\delta = \mu_T - \mu_R$, σ_T^2, σ_R^2, and ω_{RT} (the covariance of Reference and Test observations) according to the methods described in Vonesh and Chinchilli (1997) where

$$\hat{\delta}^2 \sim \left(\frac{\sigma_T^2 + \sigma_R^2 - 2\omega_{RT}}{n}\right)\chi_1^{2'}\left(\frac{\delta^2}{\frac{\sigma_T^2+\sigma_R^2-2\omega_{RT}}{n}}\right)$$

where $\chi^{2'}\left(\frac{\delta^2}{\frac{\sigma_T^2+\sigma_R^2-2\omega_{RT}}{n}}\right)$ represents the non-central chi-squared distribution with non-centrality parameter $\left(\frac{\delta^2}{\frac{\sigma_T^2+\sigma_R^2-2\omega_{RT}}{n}}\right)$. The sample variance-covariance matrix may be constructed based on the principles described by Muirhead (1982) adjusting for the cross-over nature of the study (Jones and Kenward, 1989; Vonesh and Chinchilli, 1997). Let $\hat{\boldsymbol{S}}$ be $\begin{pmatrix} \hat{\sigma}_R^2 & \hat{\omega}_{RT} \\ \hat{\omega}_{RT} & \hat{\sigma}_T^2 \end{pmatrix}$ which is distributed according to a Wishart distribution $W_2(\nu, \boldsymbol{S})$ where $\boldsymbol{S}$ is the matrix $\begin{pmatrix} \sigma_R^2 & \omega_{RT} \\ \omega_{RT} & \sigma_T^2 \end{pmatrix}$. The $E(\hat{\boldsymbol{S}}) = \boldsymbol{S}$ and $Var(\hat{\sigma}_t^2) = \frac{2\sigma_t^4}{n-2}$ for $(t = T, R)$, $Var(\hat{\omega}_{RT}) = \frac{2\omega_{RT}^2}{n-2}$, $COV(\hat{\sigma}_T^2, \hat{\sigma}_R^2) = \frac{2\omega_{RT}^2}{n-2}$, and $COV(\hat{\sigma}_t^2, \hat{\omega}_{RT}) = \frac{2\sigma_t^2\omega_{RT}}{n-2}$ for $(t = T, R)$ (Muirhead, p.90, 1982). We now derive the expected value and variance for the FDA's proposed estimate of PBE (FDA Guidance, 2001) in such a design and derive an unbiased estimator based on the findings.

Theorem 4.1 *Bias and Variance in Linearised PBE Metrics in a Two-period, Balanced Cross-over Estimated using Method-of-Moments*

When a method-of-moments estimator for the reference-scaled FDA metric is derived for (49) from a balanced two-period cross-over using a plug-in approach with method-of-moment estimates, then the expected value is

$$\left(\frac{\sigma_T^2 + \sigma_R^2 - 2\omega_{RT}}{n}\right) + \delta^2 + \sigma_T^2 - (1 + c_{FDA})\sigma_R^2 \tag{51}$$

This expression (51) is asymptotically unbiased for (49) and has variance of

$$2\Big(\frac{\sigma_T^2+\sigma_R^2-2\omega_{RT}}{n}\Big)^2+4\Big(\frac{\sigma_T^2+\sigma_R^2-2\omega_{RT}}{n}\Big)\delta^2+\frac{2\sigma_T^4}{n-2} \\ +(1+c_{FDA})^2\Big(\frac{2\sigma_R^4}{n-2}\Big)-(1+c_{FDA})\Big(\frac{2\omega_{RT}^2}{n-2}\Big) \tag{52}$$

An unbiased reference-scaled estimator for (49) is

$$\hat{\delta}^2+(\frac{n-1}{n})\hat{\sigma}_T^2-(1+c_{FDA}+\frac{1}{n})\hat{\sigma}_R^2+\frac{2\hat{\omega}_{RT}}{n} \tag{53}$$

and has variance of

$$2\Big(\frac{\sigma_T^2+\sigma_R^2-2\omega_{RT}}{n}\Big)^2+4\Big(\frac{\sigma_T^2+\sigma_R^2-2\omega_{RT}}{n}\Big)\delta^2+\frac{2(1-\frac{1}{n})^2\sigma_T^4}{n-2}$$

$$+(1+c_{FDA}+\frac{1}{n})^2\Big(\frac{2\sigma_R^4}{n-2}\Big)+\Big(\frac{8\omega_{RT}^2}{n^2(n-2)}\Big)-4(1-\frac{1}{n})(1+c_{FDA}+\frac{1}{n})\frac{\omega_{RT}^2}{n-2}$$

$$+\frac{8(1-\frac{1}{n})\sigma_T^2\omega_{RT}}{n-2}-8(1+c_{FDA}+\frac{1}{n})\Big(\frac{\sigma_R^2\omega_{RT}}{n(n-2)}\Big)$$

When a method-of-moments estimator for the constant-scaled FDA metric is derived for (50) from a balanced two-period cross-over as

$$\hat{\delta}^2+\hat{\sigma}_T^2-\hat{\sigma}_R^2-c_{FDA}(0.04)$$

then the expected value is

$$\Big(\frac{\sigma_T^2+\sigma_R^2-2\omega_{RT}}{n}\Big)+\delta^2+\sigma_T^2-\sigma_R^2-c_{FDA}(0.04) \tag{54}$$

This expression (54) is asymptotically unbiased for (49) and has variance of

$$2\Big(\frac{\sigma_T^2+\sigma_R^2-2\omega_{RT}}{n}\Big)^2+4\Big(\frac{\sigma_T^2+\sigma_R^2-2\omega_{RT}}{n}\Big)\delta^2+\frac{2\sigma_T^4}{n-2} \\ +\Big(\frac{2\sigma_R^4}{n-2}\Big)-\Big(\frac{2\omega_{RT}^2}{n-2}\Big) \tag{55}$$

An unbiased estimator for (50) is

$$\hat{\delta}^2+(1-\frac{1}{n})\hat{\sigma}_T^2-(1+\frac{1}{n})\hat{\sigma}_R^2+\frac{2\hat{\omega}_{RT}}{n}-c_{FDA}(0.04) \tag{56}$$

and has variance of

$$2\Big(\frac{\sigma_T^2+\sigma_R^2-2\omega_{RT}}{n}\Big)^2+4\Big(\frac{\sigma_T^2+\sigma_R^2-2\omega_{RT}}{n}\Big)\delta^2+\frac{2(1-\frac{1}{n})^2\sigma_T^4}{n-2}$$

$$+(1+\frac{1}{n})^2\Big(\frac{2\sigma_R^4}{n-2}\Big)+\Big(\frac{8\omega_{RT}^2}{n^2(n-2)}\Big)-4(1-\frac{1}{n^2})\frac{\omega_{RT}^2}{n-2}$$

$$+\frac{4(1-\frac{1}{n})\sigma_T^2\omega_{RT}}{n(n-2)}-4(1+\frac{1}{n})\Big(\frac{\sigma_R^2\omega_{RT}}{n(n-2)}\Big)$$

Proof: Taking expectations of a plug-in estimate using method-of-moments in a two-period, balanced cross-over study and using the results of Muirhead (p 24-25, 1982), it is seen that this expression reduces to:

$$\Big(\frac{\sigma_T^2+\sigma_R^2-2\omega_{RT}}{n}\Big)(1+\frac{\delta^2}{\frac{\sigma_T^2+\sigma_R^2-2\omega_{RT}}{n}})+\sigma_T^2-(1+c_{FDA})\sigma_{WR}^2$$

which reduces to (51). As sample size increases (51) becomes,

$$lim_{n\to\infty}\Big(\frac{\sigma_T^2+\sigma_R^2-2\omega_{RT}}{n}\Big)+\delta^2+\sigma_T^2-(1+c_{FDA})\sigma_R^2$$

$$=\delta^2+\sigma_T^2-(1+c_{FDA})\sigma_R^2$$

which is thus asymptotically unbiased. The expected value for (53) using the same approach

reduces this value.

The variance of this expression is

$$VAR(\hat{\delta}^2) + VAR(\hat{\sigma}_T^2) + (1 + c_{FDA})^2(VAR(\hat{\sigma}_R^2)) + 2COV(\hat{\delta}^2, \hat{\sigma}_T^2)$$

$$-2(1 + c_{FDA})COV(\hat{\delta}^2, \hat{\sigma}_R^2) - 2(1 + c_{FDA})COV(\hat{\sigma}_T^2, \hat{\sigma}_R^2)$$

As $\hat{\delta}$ is independent of $\hat{\boldsymbol{S}}$, the covariance terms associated with these terms are null, and we are left with,

$$VAR(\hat{\delta}^2) + VAR(\hat{\sigma}_T^2) + (1 + c_{FDA})^2(VAR(\hat{\sigma}_R^2)) - 2(1 + c_{FDA})COV(\hat{\sigma}_T^2, \hat{\sigma}_R^2)$$

which equals (52) based on the findings in text above. The variance of the unbiased estimator follows and proof for the constant-scaled metric follows similar principles to the previous proof.□□□

Consider the data presented in Table 4 of Chapter 1. The estimated moments are as follows.

Table 15: Statistics and Quantiles for AUC and Cmax based on Data presented in Table 4

Parameter	N	$\hat{\delta}$	$\hat{\sigma}_T^2$	$\hat{\sigma}_R^2$	$\hat{\omega}_{RT}$	$SE(\hat{\delta})$
AUC	45	0.097	1.706	1.785	1.547	0.094
Cmax	47	0.051	0.852	0.924	0.729	0.082

Table 16: Estimates (Variance) for Linearised FDA Metric of PBE and Unbiased FDA Metric based on AUC and Cmax Data presented in Table 4

Parameter	Lin. FDA (Var)	Unb.Lin. FDA (Var)
AUC	-3.18 (0.64)	-3.26 (1.12)
Cmax	-1.68 (0.19)	-1.72 (0.30)

Note that while point estimates for the linearised FDA metric and unbiased estimators do not differ by a great amount, the variances for the unbiased estimators are nearly twice that of the biased metrics.

Thus, a great price in terms of precision would appear to be paid for accuracy. Correction for the correlation not taken into account due to the nature of the design in the metric appears to carry an associated cost in precision. This finding is not at all unexpected when it is recalled (see Chapter 2) that variance estimates in these designs are poorly characterised. In the next section, we will explore whether this impacts inference in replicate designs.

The asymptotic bias and variance of the metric derived using a REML UN model are straightforward to calculate as with IBE metrics. Again, it is known (Searle, 1971) that REML estimates for the moments of interest $\delta, \sigma_T^2, \sigma_R^2$ are asymptotically normally distributed. The large sample variance for $\underline{\hat{\beta}}$ and $\hat{\boldsymbol{\Sigma}}$ are $(\boldsymbol{X}'\boldsymbol{\Sigma}^{-}\boldsymbol{X})^{-}$ and $-E[\frac{\partial^2 \boldsymbol{L}^{-1}}{\partial \boldsymbol{\Sigma} \partial \boldsymbol{\Sigma}'}]$, respectively with covariance $\mathbf{0}$ where $\boldsymbol{L}$ is the *log*-likelihood in expression (30). Note that in this situation, $\hat{\delta}^2$ will be normally distributed in the limit with expected value $\sigma_\delta^2 + \delta^2$ and variance $2\sigma_\delta^4 + 4\sigma_\delta^2\delta^2$ where σ_δ^2 is the usual large sample variance of δ usually estimated using the $(\boldsymbol{X}'\boldsymbol{\Sigma}^{-}\boldsymbol{X})^{-}$ matrix if $\delta = 0$ and with expected value δ^2 and variance $4\sigma_\delta^2\delta^2$ if $\delta \neq 0$. As with the IBE case, here we are most interested in situations where $\delta \neq 0$ and will concentrate on aspects relating to this situation.

We now turn to the variance estimates. These are normally distributed in the limit with variance-covariance matrix appropriate to the structure of the model. For a REML UN model, where we estimate $\begin{pmatrix} \hat{\sigma}_T^2 \\ \hat{\sigma}_R^2 \\ \hat{\omega} \end{pmatrix}$, these are normally distributed with expected value $\begin{pmatrix} \sigma_T^2 \\ \sigma_R^2 \\ \omega \end{pmatrix}$ and symmetric variance-covariance of $-E[\frac{\partial^2 \boldsymbol{L}^{-1}}{\partial \boldsymbol{\Sigma} \partial \boldsymbol{\Sigma}'}]$, the terms of which we shall denote as $\begin{pmatrix} l_T & l_{TxR} & l_{Tx\omega} \\ l_{TxR} & l_R & l_{Rx\omega} \\ l_{Tx\omega} & l_{Rx\omega} & l_\omega \end{pmatrix}$

Theorem 4.2 *Asymptotic Bias and Variance of the Linearised PBE Metrics from FDA using REML Estimation in a Two-Period Cross-over*

Let

$$\hat{\nu}_{PBE} = \hat{\delta}^2 + \hat{\sigma}_T^2 - (1 + c_{FDA})\hat{\sigma}_R^2 \tag{57}$$

be an estimate for the reference-scaled PBE metric in accordance with FDA Guidance (2001)

using a REML UN model. Then, this estimate is asymptotically normally distributed, unbiased with

$$E[\hat{\nu}_{PBE}] = \delta^2 + \sigma_T^2 - (1 + c_{FDA})\sigma_R^2$$

and has variance of

$$Var[\hat{\nu}_{PBE}] = 4\sigma_\delta^2\delta^2 + l_T + (1 + c_{FDA})^2(l_R)$$

$$-2(1 + c_{FDA})l_{TxR}$$

Let

$$\hat{\nu}_{C.PBE} = \hat{\delta}^2 + \hat{\sigma}_T^2 - \hat{\sigma}_R^2 - 0.04(c_{FDA}) \tag{58}$$

be an estimate for the constant-scaled metric in accordance with FDA Guidance (2001) using a REML UN model. Then, this estimate is asymptotically normally distributed, unbiased with

$$E[\hat{\nu}_{C.PBE}] = \delta^2 + \sigma_T^2 - \sigma_R^2 - 0.04(c_{FDA})$$

and has variance of

$$Var[\hat{\nu}_{C.PBE}] = 4\sigma_\delta^2\delta^2 + l_T + l_R - 2l_{TxR}$$

Proof: The proofs follow the general structure of theorems in Chapter 3 under the asymptotic variance-covariance structure described above and are omitted here.□□□

An asymptotically normal confidence interval for the metric can easily be computed using *SAS*© PROC MIXED code in the same manner as that described in the previous Chapter for IBE.

We now derive the moments (and unbiased estimators) for the FDA's PBE metric in replicate designs.

Theorem 4.3 *Bias and Variance in the Linearised PBE Metrics in a Balanced, Replicate Design Cross-over Estimated using Method-of-Moments*

In a replicate design, the 'plug-in' reference-scaled estimator for (49) is expressed as

$$\hat{\delta}^2 + \hat{\sigma}_{\bar{T}}^2 + \frac{\hat{\sigma}_{WT}^2}{2} - (1 + c_{FDA})\left(\hat{\sigma}_{\bar{R}}^2 + \frac{\hat{\sigma}_{WR}^2}{2}\right)$$

where $\hat{\sigma}^2_{\bar{t}}$ $(t = T, R)$ *is the estimated variance of the average of test and reference formulations respectively. This expression has expected value*

$$\frac{\sigma^2_I}{n} + \delta^2 + \sigma^2_T - (1 + c_{FDA})\sigma^2_R$$

which is asymptotically unbiased for (49) with variance

$$\left(\frac{\sigma^2_I}{n}\right)^2 (2 + \frac{4\delta^2}{\sigma^2_I/n}) + \frac{2\sigma^4_{\bar{T}}}{n-p} + \frac{\sigma^4_{WT}}{2(n-p)} + (1 + c_{FDA})^2 \frac{2\sigma^4_{\bar{R}}}{n-p} + (1 + c_{FDA})^2 \frac{\sigma^4_{WR}}{2(n-p)}$$

$$-4(1 + c_{FDA}) \frac{\omega^2_{RT}}{n-p}$$

An unbiased reference-scaled estimator may be derived as

$$\hat{\delta}^2 + (1 - \frac{1}{n})(\hat{\sigma}^2_{\bar{T}} + \frac{\hat{\sigma}^2_{WT}}{2}) - (1 + c_{FDA} + \frac{1}{n})(\hat{\sigma}^2_{\bar{R}} + \frac{\hat{\sigma}^2_{WR}}{2}) + \frac{2\hat{\omega}_{RT}}{n} \tag{59}$$

with variance

$$\left(\frac{\sigma^2_I}{n}\right)^2 (2 + \frac{4\delta^2}{\sigma^2_I/n}) + (1 - \frac{1}{n})^2 \frac{2\sigma^4_{\bar{T}}}{n-p} + (1 - \frac{1}{n})^2 \frac{\sigma^4_{WT}}{2(n-p)} + (1 + c_{FDA} + \frac{1}{n})^2 \frac{2\sigma^4_{\bar{R}}}{n-p}$$

$$+(1 + c_{FDA} + \frac{1}{n})^2 \frac{\sigma^4_{WR}}{2(n-p)} + \frac{8\omega^2_{RT}}{n^2(n-p)} - 4(1 - \frac{1}{n})(1 + c_{FDA} + \frac{1}{n}) \frac{\omega^2_{RT}}{n-p}$$

$$+8(1 - \frac{1}{n}) \frac{\sigma^2_{\bar{T}} \omega_{RT}}{n(n-p)} - 8(1 + c_{FDA} + \frac{1}{n}) \frac{\sigma^2_{\bar{R}} \omega_{RT}}{n(n-p)}$$

In a replicate design, the constant-scaled 'plug-in' estimator for (50) is expressed as

$$\hat{\delta}^2 + \hat{\sigma}^2_{\bar{T}} + \frac{\hat{\sigma}^2_{WT}}{2} - \left(\hat{\sigma}^2_{\bar{R}} + \frac{\hat{\sigma}^2_{WR}}{2}\right) - c_{FDA}(0.04)$$

where $\hat{\sigma}^2_{\bar{t}}$ $(t = T, R)$ *is the estimated variance of the average of test and reference formulations respectively. This expression has expected value*

$$\frac{\sigma^2_I}{n} + \delta^2 + \sigma^2_T - \sigma^2_R - c_{FDA}(0.04)$$

which is asymptotically unbiased for (50) with variance

$$\left(\frac{\sigma_I^2}{n}\right)^2(2+\frac{4\delta^2}{\sigma_I^2/n})+\frac{2\sigma_{\bar{T}}^4}{n-p}+\frac{\sigma_{WT}^4}{2(n-p)}+\frac{2\sigma_{\bar{R}}^4}{n-p}+\frac{\sigma_{WR}^4}{2(n-p)}$$

$$-4\frac{\omega_{RT}^2}{n-p}$$

An unbiased constant-scaled estimator may be derived as

$$\hat{\delta}^2+(1-\frac{1}{n})(\hat{\sigma}_{\bar{T}}^2+\frac{\hat{\sigma}_{WT}^2}{2})-(1+\frac{1}{n})(\hat{\sigma}_{\bar{R}}^2+\frac{\hat{\sigma}_{WR}^2}{2})+\frac{2\hat{\omega}_{RT}}{n}-c_{FDA}(0.04) \tag{60}$$

with variance

$$\left(\frac{\sigma_I^2}{n}\right)^2(2+\frac{4\delta^2}{\sigma_I^2/n})+(1-\frac{1}{n})^2\frac{2\sigma_{\bar{T}}^4}{n-p}+(1-\frac{1}{n})^2\frac{\sigma_{WT}^4}{2(n-p)}+(1+\frac{1}{n})^2\frac{2\sigma_{\bar{R}}^4}{n-p}$$

$$+(1+\frac{1}{n})^2\frac{\sigma_{WR}^4}{2(n-p)}+\frac{8\omega_{RT}^2}{n^2(n-p)}-4(1-\frac{1}{n^2})\frac{\omega_{RT}^2}{n-p}$$

$$+8(1-\frac{1}{n})\frac{\sigma_{\bar{T}}^2\omega_{RT}}{n(n-p)}-8(1+\frac{1}{n})\frac{\sigma_{\bar{R}}^2\omega_{RT}}{n(n-p)}$$

Proof: In a balanced, p-sequence, replicate design cross-over, independent unbiased method-of-moments estimators $\hat{\delta}$, $\hat{\sigma}_I^2$, $\hat{\sigma}_{WT}^2$, and $\hat{\sigma}_{WR}^2$ are derived for $\delta=\mu_T-\mu_R$, $\sigma_I^2=\sigma_D^2+\frac{\sigma_{WT}^2+\sigma_{WR}^2}{2}=\sigma_{BT}^2+\sigma_{BR}^2-2\omega_{RT}+\frac{\sigma_{WT}^2+\sigma_{WR}^2}{2}$, σ_{WT}^2, and σ_{WR}^2 according to the methods described in Vonesh and Chinchilli (1997) where

$$\hat{\delta}^2\sim\left(\frac{\sigma_I^2}{n}\right)\chi_1^{2'}\left(\frac{\delta^2}{\frac{\sigma_I^2}{n}}\right)$$

where $\chi^{2'}\left(\frac{\delta^2}{\frac{\sigma_I^2}{n}}\right)$ represents the non-central chi-squared distribution with non-centrality parameter $\left(\frac{\delta^2}{\frac{\sigma_I^2}{n}}\right)$. The sample variance-covariance matrices may be constructed based on the principles described by Muirhead (1982) adjusting for the cross-over nature of the study (Jones and Kenward, 1989; Vonesh and Chinchilli, 1997). Let $\hat{\boldsymbol{S}}_B$ be $\begin{pmatrix}\hat{\sigma}_{BR}^2+\frac{\hat{\sigma}_{WR}^2}{2} & \hat{\omega}_{RT}\\ \hat{\omega}_{RT} & \hat{\sigma}_{BT}^2+\frac{\hat{\sigma}_{WT}^2}{2}\end{pmatrix}=\begin{pmatrix}\hat{\sigma}_{\bar{R}}^2 & \hat{\omega}_{RT}\\ \hat{\omega}_{RT} & \hat{\sigma}_{\bar{T}}^2\end{pmatrix}$ which is distributed according to a Wishart distribution $W_2(n-p,\boldsymbol{S}_B)$ where $\boldsymbol{S}_B$ is the matrix $\begin{pmatrix}\sigma_{BR}^2+\frac{\sigma_{WR}^2}{2} & \omega_{RT}\\ \omega_{RT} & \sigma_{BT}^2+\frac{\sigma_{WT}^2}{2}\end{pmatrix}=\begin{pmatrix}\sigma_{\bar{R}}^2 & \omega_{RT}\\ \omega_{RT} & \sigma_{\bar{T}}^2\end{pmatrix}$. The $E(\hat{\boldsymbol{S}}_B)=\boldsymbol{S}_B$, and $Var(\hat{\sigma}_{\bar{t}}^2)=\frac{2\sigma_{\bar{t}}^4}{n-p}$ for $(t=T,R)$, $Var(\hat{\omega}_{RT})=\frac{2\omega_{RT}^2}{n-p}$, $COV(\hat{\sigma}_{\bar{T}}^2,\hat{\sigma}_{\bar{R}}^2)=\frac{2\omega_{RT}^2}{n-p}$, and $COV(\hat{\sigma}_{\bar{t}}^2,\hat{\omega}_{RT})=\frac{2\sigma_{\bar{t}}^2\omega_{RT}}{n-p}$ for $(t=T,R)$ (Muirhead, p.90, 1982).

Similarly, let $\hat{\boldsymbol{S}}_W$ be defined in accordance with Vonesh and Chinchilli (1997) such that it

is independent of $\hat{\boldsymbol{S}}_B$ (see also Theorem 2.1) and equal to $\begin{pmatrix} \hat{\sigma}^2_{WR} & 0 \\ 0 & \hat{\sigma}^2_{WT} \end{pmatrix}$ which is distributed according to a Wishart distribution $W_2(n-p, \boldsymbol{S}_W)$ where $\boldsymbol{S}_W$ is the matrix $\begin{pmatrix} \sigma^2_{WR} & 0 \\ 0 & \sigma^2_{WT} \end{pmatrix}$. The $E(\hat{\boldsymbol{S}}_W) = \boldsymbol{S}_W$, and $Var(\hat{\sigma}^2_{Wt}) = \frac{2\sigma^4_{Wt}}{n-p}$ for $(t = T, R)$ (Muirhead, p.90, 1982). Note that $\hat{\delta}^2$, $\hat{\boldsymbol{S}}_B$, and $\hat{\boldsymbol{S}}_W$ are pairwise independent (Vonesh and Chinchilli, 1997).

Now,

$$\hat{\nu}_{PBE} = \hat{\delta}^2 + (\hat{\sigma}^2_{\bar{T}} + \frac{\hat{\sigma}^2_{WT}}{2}) - (1 + c_{FDA})(\hat{\sigma}^2_{\bar{R}} + \frac{\hat{\sigma}^2_{WR}}{2})$$

and from the results above and Muirhead (p.25-27, 1982) we see that

$$E[\hat{\nu}_{PBE}] = \frac{\sigma^2_I}{n} + \delta^2 + (\sigma^2_{\bar{T}} + \frac{\sigma^2_{WT}}{2}) - (1 + c_{FDA})(\sigma^2_{\bar{R}} + \frac{\sigma^2_{WR}}{2})$$

which is obviously asymptotically unbiased as $n \to \infty$. The variance of $\hat{\nu}_{PBE}$ is

$$Var(\hat{\delta}^2) + Var(\hat{\sigma}^2_{\bar{T}}) + (1/4)Var(\hat{\sigma}^2_{WT}) + (1 + c_{FDA})^2(Var\hat{\sigma}^2_{\bar{R}})$$

$$+\frac{1}{4}(1 + c_{FDA})^2 Var(\sigma^2_{WR}) - 2(1 + c_{FDA})COV(\hat{\sigma}^2_{\bar{T}}, \hat{\sigma}^2_{\bar{R}})$$

$$= (\frac{\sigma^4_I}{n^2})\Big(2 + 4\Big(\frac{\delta^2}{\sigma^2_I}\Big)\Big) + \frac{2\sigma^4_{\bar{T}}}{n-p} + \frac{\sigma^4_{WT}}{2(n-p)}$$

$$+(1 + c_{FDA})^2 \frac{2\sigma^4_{\bar{R}}}{n-p} + (1 + c_{FDA})^2 \frac{\sigma^4_{WR}}{2(n-p)} - 4(1 + c_{FDA})\frac{\omega^2_{RT}}{n-p}$$

Since, $\sigma^2_I = (\sigma^2_{BT} + \frac{\sigma^2_{WT}}{2}) + (\sigma^2_{BR} + \frac{\sigma^2_{WR}}{2}) - 2\omega_{RT} = \sigma^2_{\bar{T}} + \sigma^2_{\bar{R}} - 2\omega_{RT}$, then $E[\hat{\nu}_{PBE}] = \delta^2 + (1 + \frac{1}{n})(\sigma^2_{\bar{T}} + \frac{\sigma^2_{WT}}{2}) - (1 + c_{FDA} - \frac{1}{n})(\sigma^2_{\bar{R}} + \frac{\sigma^2_{WR}}{2}) - \frac{2\omega_{RT}}{n}$ from which the unbiased estimator (59) for the linearised PBE metric from FDA is immediately derived. The variance of the unbiased estimator follows from the principles above. The proof for other cases involving the constant-scaled metric follow similarly.□□□

This however addresses only balanced, replicate design data sets with no missing data. We now turn to REML modelling results in replicate design studies for assessing reference and constant-scaled metrics. Estimates arising from the use of REML models are model dependent (as will be seen in the retrospective analysis section of this Chapter.) We will consider the use of an unconstrained REML procedure (UN) and as an alternative will consider a constrained procedure (CSH).

It is known (Searle, 1971) that REML estimates for the moments of interest $\delta, \sigma^2_{BT}, \sigma^2_{BR}, \sigma^2_{WT}, \sigma^2_{WR}$

are asymptotically normally distributed. The large sample variance for $\hat{\underline{\beta}}$ and $\hat{\boldsymbol{\Sigma}}$ are $(\boldsymbol{X}'\boldsymbol{\Sigma}^{-}\boldsymbol{X})^{-}$ and $-E[\frac{\partial^2 \boldsymbol{L}^{-1}}{\partial \boldsymbol{\Sigma} \partial \boldsymbol{\Sigma}'}]$, respectively with covariance $\mathbf{0}$ where $\boldsymbol{L}$ is the *log*-likelihood in expression (30). Note again that in this situation, $\hat{\delta}^2$ will be normally distributed in the limit with expected value $\sigma_\delta^2 + \delta^2$ and variance $2\sigma_\delta^4 + 4\sigma_\delta^2\delta^2$ where σ_δ^2 is the usual large sample variance of δ usually estimated using the $(\boldsymbol{X}'\boldsymbol{\Sigma}^{-}\boldsymbol{X})^{-}$ matrix if $\delta = 0$ and with expected value δ^2 and variance $4\sigma_\delta^2\delta^2$ if $\delta \neq 0$. As with the IBE case, here we are most interested in situations where $\delta \neq 0$ and will concentrate on aspects relating to this situation.

We now turn to the variance estimates. As in IBE, these are normally distributed in the limit with variance-covariance matrix appropriate to the structure of the model. $\begin{pmatrix} \hat{\sigma}^2_{BT} \\ \hat{\omega}_{RT} \\ \hat{\sigma}^2_{BR} \\ \hat{\sigma}^2_{WT} \\ \hat{\sigma}^2_{WR} \end{pmatrix}$, these are normally distributed with expected value $\begin{pmatrix} \sigma^2_{BT} \\ \omega_{RT} \\ \sigma^2_{BR} \\ \sigma^2_{WT} \\ \sigma^2_{WR} \end{pmatrix}$ and symmetric variance-covariance of $-E[\frac{\partial^2 \boldsymbol{L}^{-1}}{\partial \boldsymbol{\Sigma} \partial \boldsymbol{\Sigma}'}]$, the terms of which we shall denote as $\begin{pmatrix} l_{BT} & l_{BT\text{x}\omega} & l_{BT\text{x}BR} & l_{BT\text{x}WT} & l_{BT\text{x}WR} \\ l_{BT\text{x}\omega} & l_{\omega} & l_{BR\text{x}\omega} & l_{\omega\text{x}WT} & l_{\omega\text{x}WR} \\ l_{BT\text{x}BR} & l_{BR\text{x}\omega} & l_{BR} & l_{BR\text{x}WT} & l_{BR\text{x}WR} \\ l_{BT\text{x}WT} & l_{\omega\text{x}WT} & l_{BR\text{x}WT} & l_{WT} & l_{WT\text{x}WR} \\ l_{BT\text{x}WR} & l_{\omega\text{x}WR} & l_{BR\text{x}WR} & l_{WT\text{x}WR} & l_{WR} \end{pmatrix}$

Similarly, for the constrained REML model CSH, we find that where we estimate $\begin{pmatrix} \hat{\sigma}^2_{BT} \\ \hat{\sigma}^2_{BR} \\ \hat{\rho} \\ \hat{\sigma}^2_{WT} \\ \hat{\sigma}^2_{WR} \end{pmatrix}$, these are normally distributed with expected value $\begin{pmatrix} \sigma^2_{BT} \\ \sigma^2_{BR} \\ \rho \\ \sigma^2_{WT} \\ \sigma^2_{WR} \end{pmatrix}$ and symmetric variance-covariance of $-E[\frac{\partial^2 \boldsymbol{L}^{-1}}{\partial \boldsymbol{\Sigma} \partial \boldsymbol{\Sigma}'}]$, the terms of which we shall denote as $\begin{pmatrix} l_{BT} & l_{BT\text{x}BR} & l_{BT\text{x}\rho} & l_{BT\text{x}WT} & l_{BT_W R} \\ l_{BT\text{x}BR} & l_{BR} & l_{BR\text{x}\rho} & l_{BR\text{x}WT} & l_{BR\text{x}WR} \\ l_{BT\text{x}\rho} & l_{BR\text{x}\rho} & l_{\rho} & l_{\rho\text{x}WT} & l_{\rho\text{x}WR} \\ l_{BT\text{x}WT} & l_{BR\text{x}WT} & l_{\rho\text{x}WT} & l_{WT} & l_{WT\text{x}WR} \\ l_{BT_W R} & l_{BR\text{x}WR} & l_{\rho\text{x}WR} & l_{WT\text{x}WR} & l_{WR} \end{pmatrix}$

From these definitions, it is easy to derive the expected values and variances of the relevant estimators of the FDA metric (Serfling, 1980).

Theorem 4.4 *Asymptotic Bias and Variance of the Linearised PBE Metrics from FDA in Replicate Designs using REML Estimation*

Let

$$\hat{\nu}_{PBE} = \hat{\delta}^2 + \hat{\sigma}^2_{BT} + \hat{\sigma}^2_{WT} - (1 + c_{FDA})(\hat{\sigma}^2_{WR} + \hat{\sigma}^2_{BR}) \tag{61}$$

be an estimate for the reference-scaled metric in accordance with FDA Guidance (2001) using a

REML UN model. Then, this estimate is asymptotically normally distributed, unbiased with

$$E[\hat{\nu}_{PBE}] = \delta^2 + \sigma^2_{BT} + \sigma^2_{WT} - (1 + c_{FDA})(\sigma^2_{WR} + \sigma^2_{BR})$$

and has variance of

$$Var[\hat{\nu}_{PBE}] = 4\sigma^2_{\delta}\delta^2 + l_{BT} + l_{WT} + (1 + c_{FDA})^2(l_{BR} + l_{WR}) +$$

$$2l_{BTxWT} - 2(1 + c_{FDA})l_{BTxBR} - 2(1 + c_{FDA})l_{BTxWR} - 2(1 + c_{FDA})l_{BRxWT}$$

$$-2(1 + c_{FDA})l_{WTxWR} + 2(1 + c_{FDA})^2(l_{BRxWR})$$

Let

$$\hat{\nu}_{C.PBE} = \hat{\delta}^2 + \hat{\sigma}^2_{BT} + \hat{\sigma}^2_{WT} - (\hat{\sigma}^2_{WR} + \hat{\sigma}^2_{BR}) - 0.04(c_{FDA}) \tag{62}$$

be an estimate for the constant-scaled metric in accordance with FDA Guidance (2001) using a REML UN model. Then, this estimate is asymptotically normally distributed, unbiased with

$$E[\hat{\nu}_{C.PBE}] = \delta^2 + \sigma^2_{BT} + \sigma^2_{WT} - (\sigma^2_{WR} + \sigma^2_{BR}) - 0.04(c_{FDA})$$

and has variance of

$$Var[\hat{\nu}_{C.PBE}] = 4\sigma^2_{\delta}\delta^2 + l_{BT} + l_{WT} +$$

$$l_{BR} + l_{WR} + 2l_{BTxWT} - 2l_{BTxBR} - 2l_{BTxWR} - 2l_{BRxWT} - 2l_{WTxWR} + 2l_{BRxWR}$$

Similarly, let

$$\hat{\nu}_{PBE} = \hat{\delta}^2 + \hat{\sigma}^2_{BT} + \hat{\sigma}^2_{WT} - (1 + c_{FDA})(\hat{\sigma}^2_{WR} + \hat{\sigma}^2_{BR}) \tag{63}$$

be an estimate for the reference-scaled metric in accordance with FDA Guidance (2001) using a REML CSH model. Then, this estimate is asymptotically normally distributed, unbiased with

$$E[\hat{\nu}_{PBE}] = \delta^2 + \sigma^2_{BT} + \sigma^2_{WT} - (1 + c_{FDA})(\sigma^2_{WR} + \sigma^2_{BR})$$

and has variance of

$$Var[\hat{\nu}_{PBE}] = 4\sigma_{\delta}^2\delta^2 + l_{BT} + (1 + c_{FDA})^2(l_{BR}) + l_{WT} + (1 + c_{FDA})^2(l_{WR})$$

$$-2(1 + c_{FDA})l_{BTxBR} + 2l_{BTxWT} - 2(1 + c_{FDA})l_{BTxWR} - 2(1 + c_{FDA})l_{BRxWT} +$$

$$2(1 + c_{FDA})^2 l_{BRxWR} - 2(1 + c_{FDA})l_{WTxWR}$$

Let

$$\hat{\nu}_{C.PBE} = \hat{\delta}^2 + \hat{\sigma}_{BT}^2 + \hat{\sigma}_{WT}^2 - (\hat{\sigma}_{WR}^2 + \hat{\sigma}_{BR}^2) - 0.04(c_{FDA}) \tag{64}$$

be an estimate for the constant-scaled metric in accordance with FDA Guidance (2001) using a REML CSH model. Then, this estimate is asymptotically normally distributed, unbiased with

$$E[\hat{\nu}_{C.PBE}] = \delta^2 + \sigma_{BT}^2 + \sigma_{WT}^2 - (\sigma_{WR}^2 + \sigma_{BR}^2) - 0.04(c_{FDA})$$

and has variance of

$$Var[\hat{\nu}_{C.PBE}] = 4\sigma_{\delta}^2\delta^2 + l_{BT} + l_{WT} +$$

$$l_{BR} + l_{WR} + 2l_{BTxWT} - 2l_{BTxBR} - 2l_{BTxWR} - 2l_{BRxWT} - 2l_{WTxWR} + 2l_{BRxWR}$$

Proof: The proofs follow the general structure of those used for Chapter 3 based on the above variance-covariance matrices and are omitted here.□□□

An asymptotically normal confidence interval for the metric can easily be computed using *SAS*© PROC MIXED code in the same manner as previously described.

4.2.3 Testing and Inferential Procedures for PBE

We now consider testing and inferential procedures for PBE criteria. Practical testing and inferential procedures for PBE have received little attention in the literature. Quiroz et al. (2002) describes an asymptotic procedure using method-of-moments estimation; however, as findings are similar to those produced by the FDA Guidance (2001) procedure developed based on the findings in Hyslop et al. (2001-2002), we will include only the latter in the assessments of this chapter. Hauck et al. (1997) and Welleck (2000) also considered alternative approaches, but we will not further develop these apporaches here. Some have noted that the FDA's procedure to

be flawed in that it does not take into account the covariance produced by repeated observations on the same subject; however, as we will see in the retrospective analysis of this Chapter and the simulations of Chapter 5, little practical differences in inference are evident.

For the linearised version of the FDA's PBE metric (49), a procedure is described in the FDA Guidance (2001) that is appropriate for parallel group designs - i.e. where the mutually independent estimates $\hat{\delta}$, $\hat{\sigma}_T^2$, and $\hat{\sigma}_R^2$ are derived for $\delta = \mu_T - \mu_R$, σ_T^2, and σ_R^2 where

$$\hat{\delta}^2 \sim \left(\frac{\sigma_T^2}{n_T} + \frac{\sigma_R^2}{n_R}\right)\chi_1^{2'}\left(\frac{\delta^2}{\frac{\sigma_T^2}{n_T} + \frac{\sigma_R^2}{n_R}}\right)$$

where $\chi^{2'}\left(\frac{\delta^2}{\frac{\sigma_T^2}{n_T}+\frac{\sigma_R^2}{n_R}}\right)$ represents the non-central chi-squared distribution with non-centrality parameter $(\frac{\delta^2}{\frac{\sigma_T^2}{n_T}+\frac{\sigma_R^2}{n_R}})$,

where

$$\hat{\sigma}_T^2 \sim \frac{\sigma_T^2(\chi_{\nu_T}^2)}{\nu_T}$$

with $\chi_{\nu_T}^2$ representing the central chi-squared distribution with $\nu_T = n_T - 1$ degrees of freedom,

and

where

$$\hat{\sigma}_R^2 \sim \frac{\sigma_R^2(\chi_{\nu_R}^2)}{\nu_R}$$

with $\chi_{\nu_R}^2$ representing the central chi-squared distribution with $\nu_R = n_R - 1$ degrees of freedom. This procedure is summarised below. Here an approximate procedure such as that developed by Hyslop et al. (2000) is proposed based on the findings of Ting et al. (1990) and presented in the FDA Guidance (2001).

1. Derive unbiased, independent estimators $\hat{\delta}$, $\hat{\sigma}_T^2$, and $\hat{\sigma}_R^2$ for $\delta = \mu_T - \mu_R$, σ_T^2, and σ_R^2 as described earlier in this Chapter.

2. Let H_δ be the square of the absolute value of the larger of the lower and upper 90% bounds on δ derived using the t-distribution and using Satterthwaite approximation for the degrees of freedom, $H_T = \frac{\nu_T(\hat{\sigma}_T^2)}{\chi_{\nu_T}^2(0.05)}$, and $H_R = \frac{-(1+c_{FDA})\nu_R\hat{\sigma}_R^2}{\chi_{\nu_R}^2(0.95)}$ where $\chi_{\nu_t}^2(\alpha)$ is the αth-percentile point of the Chi-squared distribution with ν_t degrees of freedom.

3. Then

$$(\hat{\delta}^2 + \hat{\sigma}_T^2 - (1 + c_{FDA})\hat{\sigma}_R^2) \mp [(H_\delta - \hat{\delta}^2)^2 + (H_T - \hat{\sigma}_T^2)^2 + (H_R - (-(1 + c_{FDA})\hat{\sigma}_R^2))^2]^{\frac{1}{2}}$$

is an approximate, 90% confidence interval for (49).

The properties of nonparametric bootstrap-based inference, extending the results of Shao et al. (2000a-b), is a trivial exercise (data on file) and will not be reproduced here. Obviously, for the reasons discussed in Chapter 1, it is be desirable to have an approximate or asymptotic procedure in addition to this non-parametric method, and we will develop an asymptotic approach using the REML properties of the previous section similar to that developed for IBE in Chapter 3.

The findings in the previous Chapters also allow for the derivation of asymptotic REML confidence intervals for the linearised FDA criterion in a straightforward manner (see Chapter 3).

Confidence intervals constructed based on the REML normal asymptotic approximation (described above), the Cornish-Fisher approximation, and using the non-parametric percentile method (2000 bootstrap samples maintaining the number of subjects per sequence) for Linearised PBE FDA metrics are below based on the data of Table 4. Inference is the same using all three approaches (PBE is demonstrated for AUC and Cmax). Note that the normal approximation falls between the bootstrapped bound and the upper bound constructed using the Cornish-Fisher approximation. The non-parametric percentile bootstrap is known to be conservative in Type I error rate and the Cornish-Fisher approximation is known to be anti-conservative with respect to maintenance of the Type I error rate.

Table 17: Upper 95% Bounds for the Linearised FDA Metric of PBE based on AUC and Cmax Data presented in Table 4

Parameter	Normal Appx	CF Appx	NP PCL
AUC	-1.87	-1.60	-1.93
Cmax	-0.97	-0.89	-1.16

4.3 Retrospective Analysis

In the retrospective analysis of the replicate design data sets, we will consider:

1. Bias in the estimates arising from Method-of-Moments estimation

2. PBE inference reached based on comparison of the bootstrap, Hyslop's procedure, and the application of asymptotically normal confidence intervals from REML.

The same data was utilised as was in Chapter 2-3.

We now turn to discussion of retrospective analysis of the replicate design data sets for the assessment of population bioequivalence. Three inferential methods were explored:

1. The approximation-method based on the Cornish-Fisher expansion was applied to the data sets in accordance with FDA Guidance (2001) using method-of-moments estimates to derive the upper bound of a 90% confidence interval for the linearised FDA population bioequivalence metric.

2. The non-parametric percentile method (Efron and Tibshirani, 1993) was used with 2000 bootstraps, maintaining the observed numbers of subjects per sequence, and using an unrestricted (UN) REML model to derive the upper bound of a 90% confidence interval for the linearised FDA population bioequivalence metric. The unrestricted model was selected instead of a restricted REML model (recommended in FDA Guidance, 1997) in order to provide consistency relative to the method-of-moments procedure used in the Hyslop et al (2000) method described above. Method-of-moments and unrestricted REML estimates should be equal when the data set is strongly balanced (Vonesh and Chinchilli, 1997) and has no missing data. Also this model was selected in order to provide consistency with the asymptotic procedure developed in previously in this Chapter.

3. The asymptotic procedure developed previously in this Chapter was applied to each data set using an unrestricted REML model and a restricted REML model.

We first describe the results of each analysis and then compare and contrast between them. This sub-section ends with discussion of significant findings to be explored through the use of simulation and conclusions.

When Hyslop et al's (2000) approximate procedure was applied to each data set for the assessment of population bioequivalence, it was found that three AUC data sets (Q1, X, Y) and five Cmax data sets (E, G, N1, X, ZB) failed to demonstrate bioequivalence with upper bounds

falling above zero (see Table 54 and 55). However, the non-parametric percentile bootstrap method found that one AUC data set (N1) and five Cmax data sets (G, N1, P, Y, ZB) failed to demonstrate PBE (see Tables 56 and 57). When assessed using the unrestricted REML asymptotic procedure (Tables 56 and 60), no AUC data sets failed to demonstrate PBE, and only three Cmax data sets (G, N1, and ZB) failed to demonstrate PBE.

Thus in the vast majority of cases, all three methods of inference indicate that most data sets demonstrated population bioequivalence. Of those that failed, for Cmax, data sets G, N1, and ZB are consistently identified by all procedures as failing to demonstrate PBE.

The results for data set N1 (AUC) and P (Cmax) again illustrated (for previous discussion see Zariffa et al., 2000) the peculiar behaviour of the FDA criterion near the cut-off value for reference scaling of $\sigma_R^2 = 0.04$. Here we observed estimated $\hat{\sigma}_R = 0.204$ for N1 (AUC) and $\hat{\sigma}_R = 0.232$ for P (Cmax) just above the cut-off. It is known in instances such as this (Zariffa et al., 2000; Shao et al., 2000a-b) that the bootstrap procedure can result in inconsistent inference, and the observation that this data set fails to demonstrate PBE should therefore be interpreted with caution. Indeed previous reports (Zariffa et al., 2000) did not observe the data sets to fail PBE. Thus we believe these findings to be related to inconsistency of the bootstrap and the change to a linearised procedure.

Indeed, when Shao et al.'s (2000a-b) proposed correction to the bootstrap procedure of FDA Guidance (1997) was applied an upper bound of 1.33 for N1 (AUC) and of 0.47 for P (Cmax) results for the non-linearised criterion (values are assessed relative to a cut-off of 1.74 with value less than this denoting demonstration of PBE). Shao et al.'s proposed procedure was carried out without regard to the fact that the estimate from such an exercise is positively biased (see previous results in this Chapter). Thus we believe the findings for data set N1 (AUC) and P (Cmax) to be an artifact of the bootstrap inferential procedure in accordance with previous findings for other data sets (Zariffa et al., 2000).

Discrepancies in findings for AUC data sets Q1, X, and Y and Cmax data sets E and X were potentially more readily interpretable. Accounting for non-null correlation between variance estimates allowed the asymptotic and the non-parametric percentile bootstrap procedures to conclude PBE was demonstrated while ignoring this correlation resulted in rejection under the procedure developed by Hyslop et al. (2000). Given the number of changes, however, the

assumption of independence of variance estimates is of limited potential, and certainly not great, concern in data from cross-over studies.

The discrepancy in findings between the bootstrap and asymptotic procedures for data set Y (Cmax) did not appear to be due to any striking finding in the data set. Thus we suspect that random-error may be involved and will investigate using simulation to elicit comparative error rates.

Turning to consideration of constraints in REML modelling and its impact on inference for PBE using the asymptotic procedure, for PBE, no discrepancies in inference for AUC and Cmax were observed when constraining variance estimates to be non-zero.

Lastly, we consider the estimated bias in the population bioequivalence metric, equal as previously established in this Chapter as σ_I^2/n. We neglect that this bias is exact for only strongly balanced, complete data sets for the purposes of this exercise. The mean (STD) bias for AUC and Cmax were estimated as 0.002 (0.001) and 0.004 (0.003), respectively. Expressed as an absolute percentage of the estimate for the PBE metric, this mean (STD) was 0.55% (0.49%) and 1.62% (3.53%), respectively. Accounting for this positive bias was not found to impact inference relative to the Hyslop upper bound in any data set, as would be expected given the asymptotic unbiasedness of the metric previously established in this Chapter relative to the sample sizes employed in these studies.

Obviously, PBE allows for much easier market access than does the corresponding ABE criteria. Indeed, the vast majority of data sets demonstrate PBE regardless of inferential method. This was concerning in particular for data sets I1, I2, and T (AUC) and for data sets I1, I2, L2, Q2 and T (Cmax). These data sets demonstrated 20 to 40% changes in mean response but would still be allowed market entry under PBE due to a combination of decreased total variability in the test product and the impact of scaling to reference product variation. The widening of goalposts while indicative of carrying out a smaller study (see FDA Guidance, 2001), for such tradeoffs between moments of interest in the estimated metric for PBE is of concern when considering the potential for therapeutic failure in the marketplace.

4.4 Findings and Discussion

The procedure developed in FDA Guidance (2001) based upon Hyslop et al. (2000) for the assessment of PBE is flawed in that it does not account for the covariance among variance estimates introduced by repeated observations within-subject in cross-over designs. However, when this correlation is taken into account (using the asymptotic REML approach developed in this thesis, which should also guard against bias introduced by missing data or imbalance), little change in inference is observed. Thus the criticism leveled at the procedure would not appear to be warranted at this time based upon empirical review of the data. However, we will study whether this holds true in other situations using simulation in Chapter 5.

However, we note one very concerning finding here. It appears very easy to demonstrate PBE in the typical replicate design BE study. Several studies would permit products with large changes in average bioavailability to market. This is due to decreases in the test products variance and due to the scaling of the metric due to large total variance in the reference product. Such a test product has the potential to fail users upon marketplace entry due to failure of efficacy or the presence of intolerable side effects.

The magnitude of this potential for marketplace failure is a characteristic of the therapeutic index of the individual drug product, and hence we will not pursue the subject further here, except to note it is worrisome. Revision of the goalpost would appear to be a first step in correcting this deficiency, and we believe the goalpost chosen by FDA should be revisited if PBE is considered as an approach for market access in future (FDA Guidance, 2002).

Behaviour of the PBE criteria should be assessed cautiously when examining products with reference product total variation near the FDA cut-off of 0.04 if using a nonparametric bootstrap procedure. Consistency of such a procedure is concerning, along with its potential for Type I or II error, and we will study a correction (suggested by Shao et al., 2000a-b) to study its effects in Chapter 5.

In conclusion, the asymptotic procedure proposed in this thesis appears to result in consistent inference with alternative procedures currently available. In sample sizes usually used for the assessment of average bioequivalence in replicate design studies, bias introduced by a 'plug-in' method does not appear to be of a magnitude sufficient for concern with regard to impact on inference in balanced data sets with complete data. Inconsistency of the bootstrap procedure and

issues in correlation with the definition of FDA Guidance (2001) procedure appear to make the asymptotic assessment procedure developed in this thesis a potential alternative when reviewing the results of PBE studies.

5 Simulations and the Transitivity of Individual Bioequivalence Testing

Aspects of the findings of this chapter were presented at the American Association of Pharmaceutical Scientists joint workshop with the USA Food and Drug Administration (Zariffa and Patterson, 2000), at the American Statistical Society Joint Statistical meetings (Patterson and Jones, 2002e), and at the International Society of Clinical Biostatistics meeting (Patterson and Jones, 2002f). Aspects of the findings were published in the *Journal of Clinical Pharmacology* (Zariffa and Patterson, 2001), in *Pharmaceutical Statistics* (Patterson and Jones, 2002g), in the *Proceedings of the Joint Statistical Meetings* (Patterson and Jones, 2002h), and in a series of a GlaxoSmithKline technical reports (Patterson et al., 2001e; Patterson and Jones, 2002b-c).

5.1 Introduction and Goals of Chapter

In Chapters 2 through 4, retrospective analyses were used to assess the performance of the average bioequivalence (ABE), individual bioequivalence (IBE), and population bioequivalence (PBE) methods in practice. To briefly summarise, the FDA's recommended methods of analysis were first used to assess ABE, IBE, and PBE. A constrained REML model (option 'FA0(2)' to constrain variance estimates to be positive) was used to construct the ABE two-one sided tests as described in Chapter 2, and the Cornish-Fisher expansion (see Hyslop et al., 2000 and FDA Guidance, 2001) was used utilising method-of-moments estimates to test for IBE and PBE respectively for both AUC and Cmax. For AUC, seven data sets failed to demonstrate ABE; nine data sets failed to demonstrate IBE, and three data sets failed to demonstrate PBE using these approaches to analysis. For Cmax, 16 data sets failed to demonstrate ABE; 14 data sets failed to demonstrate IBE, and five data sets failed to demonstrate PBE. The agreement within data sets across procedures is described in Table 18.

Table 18: ABE, IBE, and PBE Findings from Retrospective Analysis of 51 Replicate Design Data Sets using the FDA Guidance (2001) Procedures

Number of Data Sets with Finding	AUC	Cmax
Fails ABE, Fails IBE, Fails PBE	0	2
Fails ABE, Pass IBE, Fails PBE	0	0
Fails ABE, Fails IBE, Pass PBE	4	5
Fails ABE, Pass IBE, Pass PBE	3	9
Pass ABE, Fails IBE, Fails PBE	0	3
Pass ABE, Pass IBE, Fails PBE	3	0
Pass ABE, Fails IBE, Pass PBE	5	4
Pass ABE, Pass IBE, Pass PBE	36	28
Total	51	51

From these findings, we can infer that IBE is less stringent than ABE in some situations. For example, in one instance, an 18% change in δ was observed (failing ABE) but passed under IBE. The magnitude and variance of the variance estimates (ill-characterised in samples this small) are the contributing factor to this in the confidence interval for (24) (as we will illustrate using simulation in this chapter). Note also, that it appears very easy to demonstrate PBE in the typical replicate design BE study for a highly variable product. Some data sets demonstrated 20 to 40% changes in mean response (failing ABE) but would still be admitted to market under the PBE approach.

Additionally findings indicate that inference may be model dependent (see Chapter 2). When using unbiased method-of-moments estimates to construct the IBE and PBE metrics in replicate cross-over designs, it is known (Chapter 3 and 4) that the estimates for (24) and (23) are positively biased by an amount proportional to $\frac{\sigma_D^2+(\frac{\sigma_{WT}^2+\sigma_{WR}^2}{2})}{n}$ in small samples and are asymptotically unbiased. In the analysis of these 51 replicate design data sets, as an absolute percentage of the estimate for the IBE metric, the positive bias mean (sd) was estimated to be 1.50% (4.61%) and 2.19% (3.83%) for AUC and Cmax, respectively. As an absolute percentage of the estimate for the PBE metric, this mean (sd) is 0.55% (0.49%) and 1.62% (3.53%) for AUC and Cmax, respectively. Such a small bias would not be expected to impact inference (given the relatively large sample sizes employed in these studies).

Retrospective comparison of method-of-moment, unconstrained and constrained REML estimation as applied to ABE may be found in Chapter 2. We now consider the agreement in inference between the Cornish-Fisher expansion (Hyslop et al., 2000 and FDA Guidance, 2001), the large sample REML testing procedures developed in Chapters 3 and 4, and a nonparametric

bootstrap procedure (FDA Guidance, 1997) using unconstrained REML estimates as applied to IBE and PBE testing. Using the asymptotic procedures for AUC, 11 data sets failed to demonstrate IBE, and no data set failed to demonstrate PBE. Using the nonparametric percentile bootstrap for AUC, 11 data sets failed to demonstrate IBE, and one data set failed to demonstrate PBE. Using the asymptotic procedure for Cmax, 16 data sets failed to demonstrate IBE while three data sets failed to demonstrate PBE. Using the nonparametric percentile bootstrap procedure for Cmax, 17 data sets failed to demonstrate IBE while five data sets failed to demonstrate PBE. The agreement across procedures is further described in the following Table 19.

Table 19: IBE Cornish-Fisher (CF), Large Sample Asymptotic (Asy), and Nonparametric Percentile Bootstrap (NP) Findings from Retrospective Analysis of 51 Replicate Design Data Sets

Number of Data Sets with Finding	AUC	Cmax
IBE		
Fails CF, Fails Asy, Fails NP	8	12
Fails CF, Pass Asy, Fails NP	0	0
Fails CF, Fails Asy, Pass NP	1	1
Fails CF, Pass Asy, Pass NP	0	1
Pass CF, Fails Asy, Fails NP	1	2
Pass CF, Pass Asy, Fails NP	2	3
Pass CF, Fails Asy, Pass NP	1	1
Pass CF, Pass Asy, Pass NP	38	31
Total	51	51
PBE		
Fails CF, Fails Asy, Fails NP	0	3
Fails CF, Pass Asy, Fails NP	0	0
Fails CF, Fails Asy, Pass NP	0	0
Fails CF, Pass Asy, Pass NP	3	2
Pass CF, Fails Asy, Fails NP	0	0
Pass CF, Pass Asy, Fails NP	1	2
Pass CF, Fails Asy, Pass NP	0	0
Pass CF, Pass Asy, Pass NP	47	44
Total	51	51

Findings suggest that the procedures are not remarkably dissimilar in inference; however, there are some discrepancies which may be associated with the potential bias in the method chosen for estimation (method-of-moments or REML) and-or Type I and II error rate. This chapter will address the questions raised by the retrospective analysis exercises undertaken in Chapters 2 through 4 with regard to the properties of ABE, IBE, and PBE metrics and inference in replicate designs. Specifically, for ABE, we will use simulation studies to assess:

1. Do estimates from the REML models and Method-of-Moments behave as normal variables

in small samples and when there is missing data?

2. What is the bias in REML estimates for the components of interest in small samples and when there is missing data?

3. Of those procedures found to provide acceptable estimates for the moments of interest, what is the Type I error rate using REML procedures in small samples?

For IBE and PBE, simulation studies will be used to assess:

1. How do the Type I error rates of the Hyslop et al. (2000), asymptotic, and bootstrap procedures compare?

2. What is the impact of missing data, and is the assumption of pairwise independence of 'plug-in' terms making up the metric critical relevance in IBE and PBE assessment?

3. Does the assumption that $\delta \neq 0$ impact IBE and PBE inference when using an asymptotic procedure?

Last, we will use simulation to assess the potential for 'bio-creep' regarding allowable changes in average exposure observed as multiple studies are used to allow multiple entries to market using IBE.

5.2 Models and Theory

Two sequence (RTRT, TRTR), replicate deigns were simulated using $SAS^{©}$ using the parameter space described in the below table. Simulations were conducted for sample sizes of 16, 24, 34, and 80 in accord with the recommended sample sizes in the FDA Guidance (2001) with equal numbers of subjects in each sequence $n/2$. Each simulation study was composed of 1000 runs. The parameter space studied is defined in the below Table 20.

Table 20: True Values used in Simulation Experiments 1 through 54 (1000 runs per simulation)

Sim	δ	σ^2_D	σ^2_{WT}	σ^2_{WR}
1	0	0	0.01	0.0025
2	0	0.001875	0.01	0.0025
3	0	0	0.01	0.01
4	0	0.0075	0.01	0.01
5	0	0	0.01	0.0225
6	0	0.016875	0.01	0.0225
7	0	0	0.09	0.04
8	0	0.03	0.09	0.04
9	0	0	0.09	0.09
10	0	0.0675	0.09	0.09
11	0	0	0.09	0.16

Table 20: True Values used in Simulation Experiments 1 through 54 (1000 runs per simulation)

Sim	δ	σ^2_D	σ^2_{WT}	σ^2_{WR}
12	0	0.12	0.09	0.16
13	0	0	0.25	0.16
14	0	0.12	0.25	0.16
15	0	0	0.25	0.25
16	0	0.1875	0.25	0.25
17	0	0	0.25	0.49
18	0	0.3675	0.25	0.49
19	0.2231	0	0.01	0.0025
20	0.2231	0.001875	0.01	0.0025
21	0.2231	0	0.01	0.01
22	0.2231	0.0075	0.01	0.01
23	0.2231	0	0.01	0.0225
24	0.2231	0.016875	0.01	0.0225
25	0.2231	0	0.09	0.04
26	0.2231	0.03	0.09	0.04
27	0.2231	0	0.09	0.09
28	0.2231	0.0675	0.09	0.09
29	0.2231	0	0.09	0.16
30	0.2231	0.12	0.09	0.16
31	0.2231	0	0.25	0.16
32	0.2231	0.12	0.25	0.16
33	0.2231	0	0.25	0.25
34	0.2231	0.1875	0.25	0.25
35	0.2231	0	0.25	0.49
36	0.2231	0.3675	0.25	0.49
37	0.6931	0	0.01	0.0025
38	0.6931	0.001875	0.01	0.0025
39	0.6931	0	0.01	0.01
40	0.6931	0.0075	0.01	0.01
41	0.6931	0	0.01	0.0225
42	0.6931	0.016875	0.01	0.0225
43	0.6931	0	0.09	0.04
44	0.6931	0.03	0.09	0.04
45	0.6931	0	0.09	0.09
46	0.6931	0.0675	0.09	0.09
47	0.6931	0	0.09	0.16
48	0.6931	0.12	0.09	0.16
49	0.6931	0	0.25	0.16
50	0.6931	0.12	0.25	0.16
51	0.6931	0	0.25	0.25
52	0.6931	0.1875	0.25	0.25
53	0.6931	0	0.25	0.49
54	0.6931	0.3675	0.25	0.49

Two scenarios were investigated. The first scenario involved studies with no missing data. In the second scenario, twelve subjects were randomly selected to have missing observations. Three subjects had a missing observation for the test formulation. Three additional subjects were missing both observations for the test formulation. Three subjects were missing an observation for both the test and reference formulations, and three subjects were missing both observations for the reference formulations. Twelve subjects were selected as this number is indicative (as a practical matter) of a setting where one would be concerned about the amount and extent (and reasons why) there was such a significant amount of missing data.

The second set of simulations were performed in order to compare the Type I error rates of the IBE procedures again utilising samples sizes of $n = 16, 24, 34$ and 80 in a two sequence, replicate design with no missing data.

Table 21: True Values used in Simulation Experiments 55 through 59 (1000 runs per simulation)

Sim	δ	σ^2_D	σ^2_{WT}	σ^2_{WR}
55	ln 1.31	0.0075	0.03	0.01
56	ln 1.25	0.03	0.06	0.04
57	ln 1.45	0.0675	0.11	0.09
58	ln 1.66	0.12	0.18	0.16
59	ln 2.49	0.3675	0.51	0.49

The third set of simulations were performed in order to compare the Type I error rates of the PBE procedures again utilising samples sizes of $n = 16, 24, 34$ and 80 in a two sequence, replicate design with no missing data.

Table 22: True Values used in Simulation Experiments 60 through 62 (1000 runs per simulation)

Sim	δ	σ^2_{BT}	σ^2_{BR}	σ^2_{WT}	σ^2_{WR}
60	ln 1.30	0.06	0.04	0.01	0.01
61	ln 1.39	0.08	0.06	0.015	0.015
62	ln 1.48	0.10	0.08	0.02	0.02

Simulations were performed using the $SAS©$ procedure 'rannor' in a manner appropriate to a replicate design and were performed using $SAS©$ version 8.1 running under UNIX. Sequence and period effects were induced in the model and were assumed to be non-null. Method-of-Moments estimation, unconstrained (UN) and Constrained (CSH, FA0(2), and RIS) $SAS©$-based REML mixed modelling procedures were conducted in accordance with the descriptions

in Chapters 2 through 4 to estimate the effects of interest in ABE, IBE, and PBE testing and will not be reproduced here. ABE testing was conducted in accordance with the descriptions of Chapter 2. For IBE and PBE, the procedure described by Hyslop et al. (2000) and in the FDA Guidance (2001) were used to derive an upper bound for the linearised FDA metrics, respectively, based on method-of-moment estimates. Asymptotic testing for IBE and PBE were conducted in accordance with the findings of Chapters 3 and 4, respectively.

Nonparametric percentile bootstrap testing for IBE and PBE were conducted in accordance with the findings of Shao et al. (2000a-b) to assess the potential of this method for correction of the previous findings of inconsistency of the bootstrap testing procedure (Zariffa et al., 2000) at the constant-scaling boundary of $\sigma^2_{WR} = 0.04$. Here 2000 bootstrap samples (subject being the bootstrapped unit) were taken (FDA Guidance, 1997). Each bootstrap sample was modelled using the unconstrained REML procedure of Chapter 2, and the linearised metric was scaled based on the magnitude of the sample's reference product variation (a slight modification to Shao et al., 2000a-b, approach based on the findings of Chapters 3-4). The upper 95th quantile serves as the upper bound of interest (Efron and Tibshirani, 1993).

We now turn to the findings from the simulation exercise. We first address the issues of bias in estimates of δ, σ^2_D, and the within-subject variance estimates. This will be followed by discussion of findings relating to Type I and II error rates in ABE, IBE, and PBE testing.

Tables of bias (SE) for δ (Tables 62, 67, 72, 77), σ^2_D (Tables 63, 68, 73, 78), σ^2_{WT} (Tables 64, 69, 74, 79), σ^2_{WR} (Tables 65, 70, 75, 80), and the FDA IBE (Tables 82, 84, 86, 88) and PBE (Tables 83, 85, 87, 89) metrics' estimates in the simulation studies may be found in the Appendix following this thesis. Findings are summarised below.

5.3 Summary of Simulation Findings for Estimation in a Replicate Cross-over Design

5.3.1 Estimation of δ

Estimates of δ were mean unbiased in all (method-of-moments, constrained and unconstrained REML) procedures in complete data sets.

When substantial missing data was introduced, REML procedures remained mean unbiased. Method-of-moments estimates were positively biased by mean of $0.05 - 0.06$ when $n = 16$,

$0.035 - 0.04$ when $n = 24$, 0.03 when $n = 34$, and 0.012 when $n = 80$.

As the Kolmogorov's test is known to be sample size dependent (Schulman, 1992; page 92), the graphical approach advocated in Fisher and van Belle (1993, page 97-100) using normal-probability plots was used to assess normality. Examples are included as follows for simulation study 1 using sample sizes of $n = 16$ in complete data sets and those with substantial missing data.

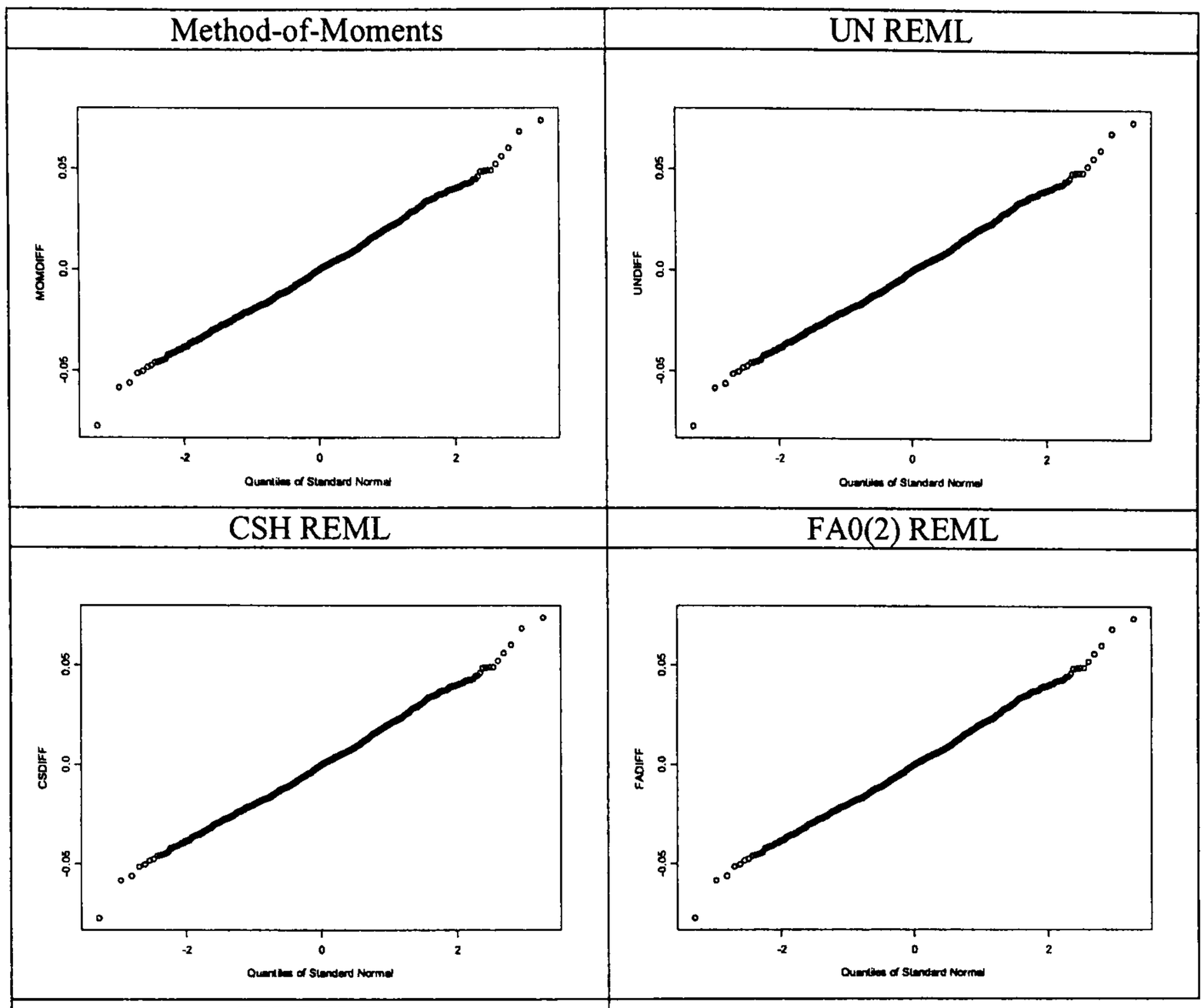


Figure 21: Normal Probability plots for method-of-moments and REML $\hat{\delta}$ from Simulation Study 1 where $n = 16$ with No Missing data

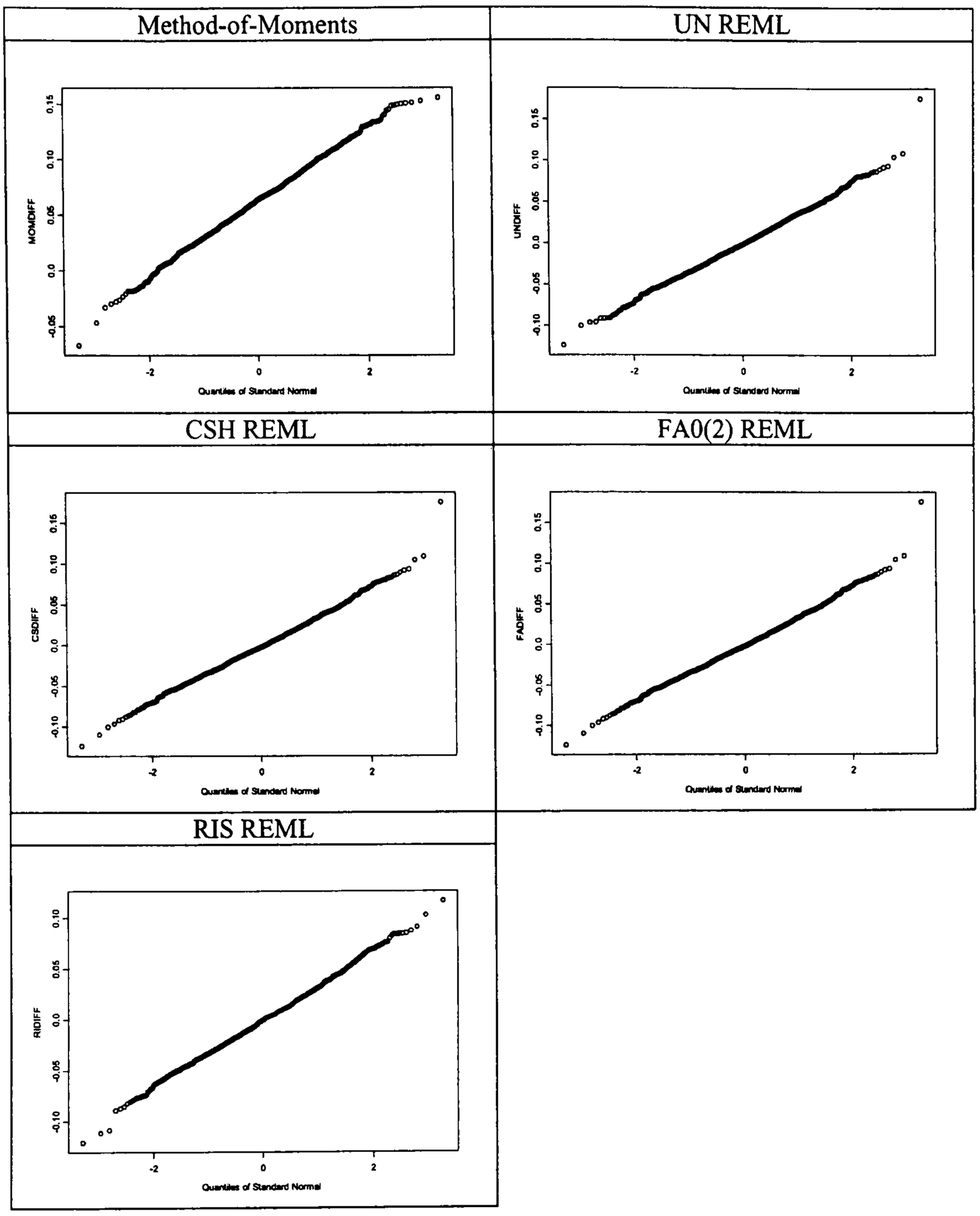


Figure 22: Normal Probability plots for method-of-moments and REML $\hat{\delta}$ from Simulation Study 1 where $n = 16$ with Substantial Missing data

Thus we confirm based on these (and other plots on file) that estimates for δ from REML and method-of-moments procedures are normally distributed and mean unbiased in complete data sets and for REML procedures with substantial missing data. While estimates were biased for the method-of-moments procedure in data sets with missing data, the estimates still appeared to behave as normally distributed variables.

The bias observed in method-of-moments estimates of δ when missing data were present are unlikely to be of concern in ABE testing. It is common practice (see FDA Guidance, 2001) to plan for deviations in δ of up to 0.05 when planning a trial. Moreover, the FDA Guidance (2001) recommends the use of a REML procedure for ABE assessment, which we know to be asymptotically unbiased (Chapter 2) and observed in this report via simulation is in general unbiased in distribution for δ in small samples.

The method-of-moments bias is more of immediate concern in PBE and IBE assessment under the recommended (FDA Guidance, 2001) approach to analysis. Chapters 3 and 4 demonstrated that the statistics used to assess PBE and IBE are positively biased in small samples when 'plugged-in' estimates of δ are unbiased. Additional bias due to the choice of method-of-moments estimation procedure in the presence of missing data for δ is of immediate concern to IBE and PBE inference.

5.3.2 Estimation of σ_D^2

We begin with discussion of the properties of σ_D^2 as a surrogate marker for switchability in individual bioequivalence.

Examination of method-of-moments σ_D^2 in the database (Figure 23) reveals the spread or variability in $\hat{\sigma}_D^2$ increases with increasing within-subject variability. This is consistent with previous simulations results (Endrenyi and Tothfalusi, 1999) and gives rise to complications in the definition of what may constitute a signal for an excessive σ_D^2 estimate. A publication by Hauck et al. (2000) makes a first attempt to address such considerations.

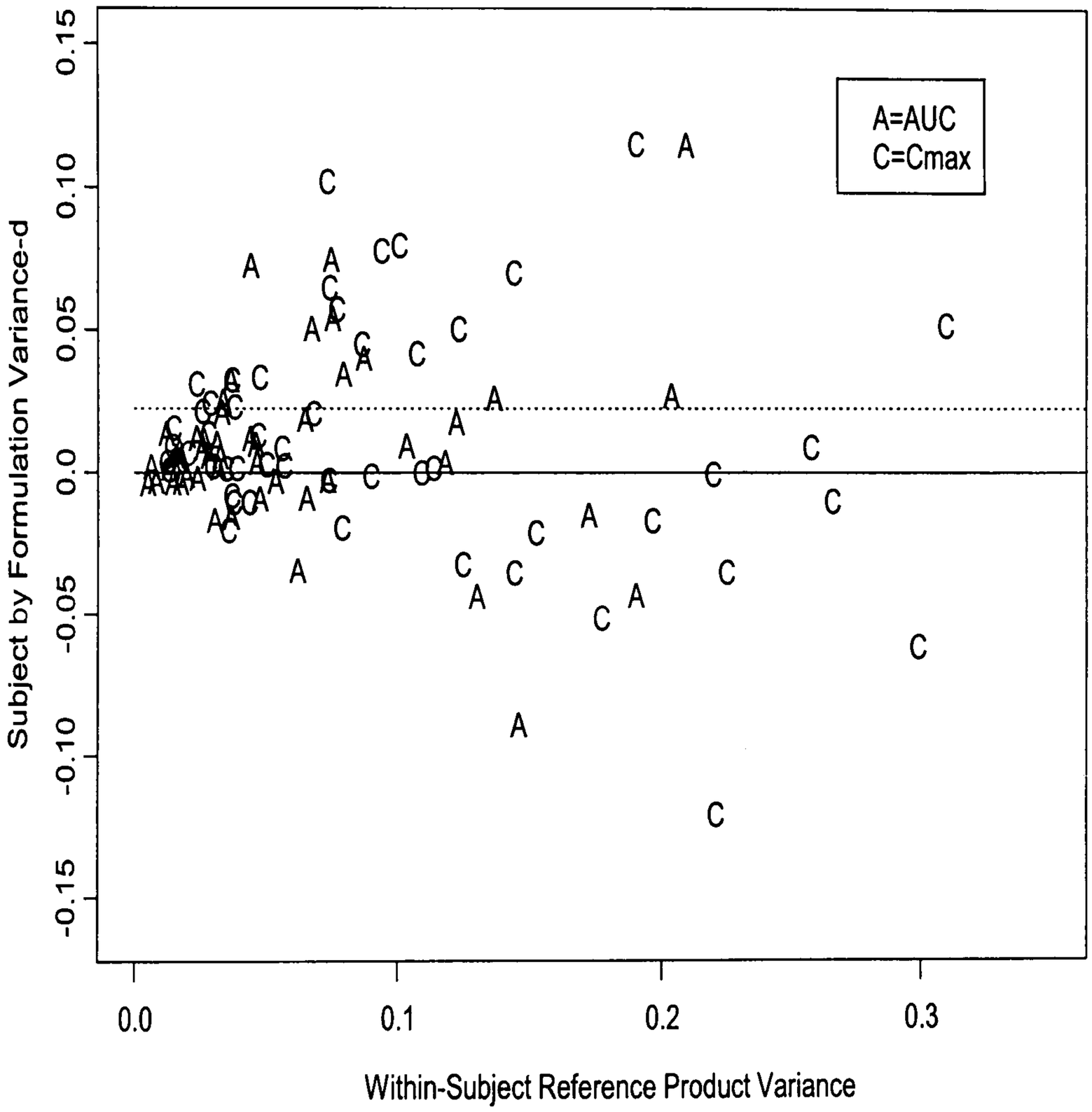


Figure 23: $\hat{\sigma}^2_D$ versus $\hat{\sigma}^2_{WR}$ in Database of 51 Replicate Designs Estimated using Method-of-Moments

Note also that negative variance estimates for σ_D^2 are not unusual when using the Method-of-Moments procedure (Figure 23). Interpretation of such negative estimates for a null or truly positive quantity is problematic, and a satisfactory resolution has never been reached in extensive statistical debates. As an alternative, the constrained (CSH) REML methodology (see FDA Guidance, 1997), gives rise to a small but detectable positive bias in σ_D^2 (Zariffa et al., 2000; Endrenyi and Tothfalusi, 1999) as between- and within-subject variance estimates and resulting σ_D^2 (see Chapter 2) are constrained to be non-negative. Other REML alternatives (see Chapter 2) have not been studied. The data clearly indicate that further evaluation of the relationship between inherent variation and the subject-by-formulation interaction (expressed as σ_D or σ_D^2) is warranted.

This factor, sometimes known as Subject-by-formulation interaction, was highlighted as a potential issue in bioequivalence studies by Ekbohm and Melander (1989). Variation associated with this interaction can be driven by a variety of factors as discussed in Chapter 1. The term σ_D^2 (and its related metric of σ_D) in the proposed aggregate IBE criterion is meant to quantify the subject-by-formulation interaction and has been the subject of intense debate since it was originally proposed in draft FDA regulatory guidance (1997). Determination of how well the measure performs may help elucidate whether it has the potential to be a meaningful surrogate marker for failure in switching products on the market. The reader is referred to Chapter 1 for a mathematical definition of σ_D^2. Of note for this thesis, σ_D^2 can be generated by differences in between-subject variation between test and reference formulations, by low correlation (ρ) and/or by a subgroup-by-formulation interaction (Hauck et al., 2000).

Initially, simulation studies were conducted using bi-variate Normal density distributions to explore the interactions between model parameters of interest. One hundred simulated studies were conducted in each run. Data sets from each simulation were analyzed according to the statistical methodology described earlier in this thesis.

It is known (Patterson et al., 1999; Zariffa et al., 2000) that variance estimates generated in bioequivalence studies powered for average bioequivalence are poorly (i.e. imprecisely) characterized, and estimates in excess of the $\sigma_D > 0.15$ cut-off (Hauck et al., 2000) should be expected due to random chance (as simulated in Figure 24). Increasing sample size does appear to provide some benefit in making these estimates quantitatively more precise (Figure 24) but the larger

sample sizes needed to achieve this are not currently recommended for demonstration of average bioequivalence in moderate variability compounds (FDA Guidance, 2001).

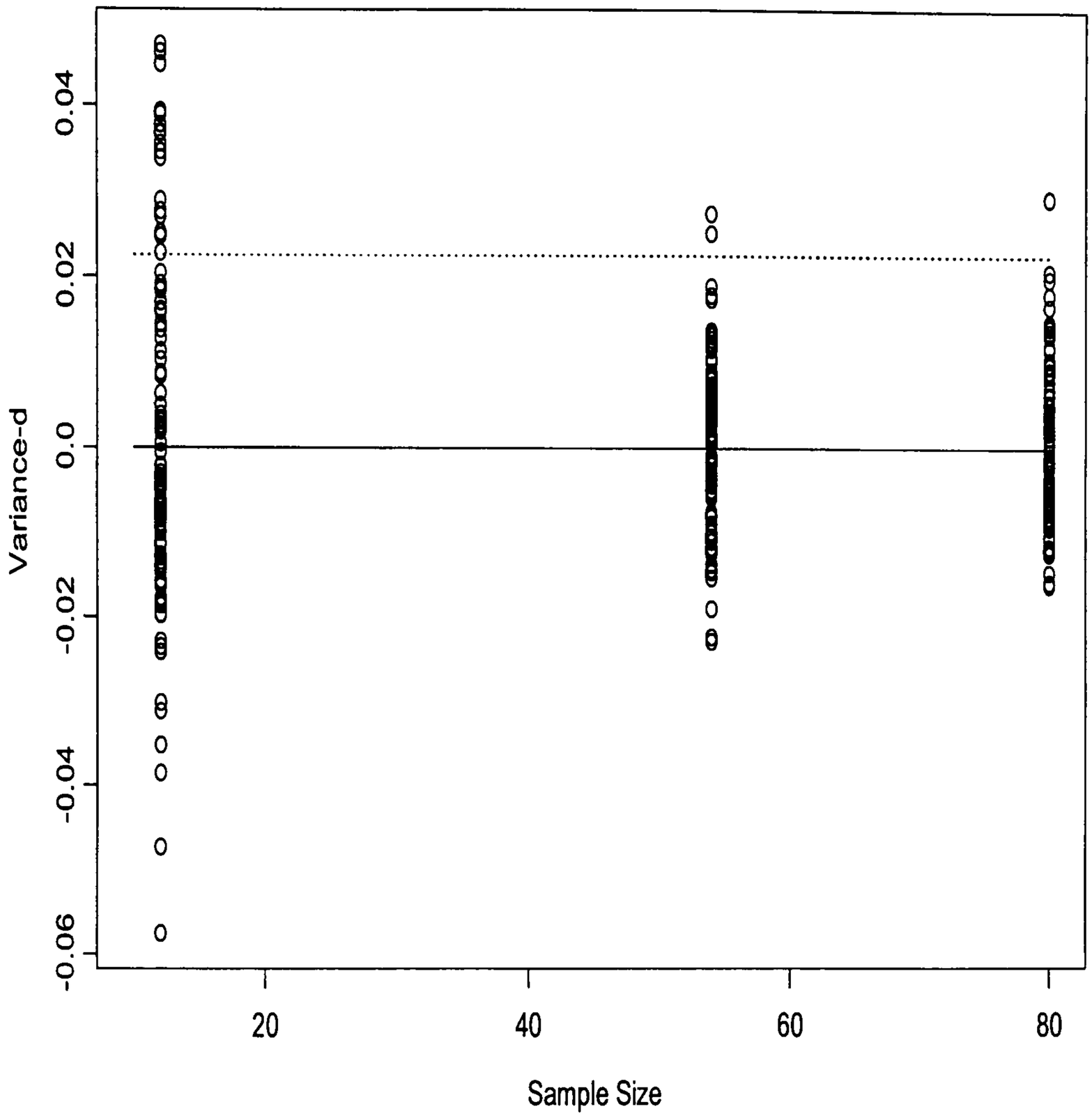


Figure 24: Simulated Method of Moments estimates for $\hat{\sigma}^2_D$ (where true $\hat{\sigma}^2_D = 0$) in Replicate Cross-over Studies as a function of sample size (n=12, 54, or 80 subjects per simulated study) for a Moderate Variability ($\hat{\sigma}^2_{WR}$=0.04) compound (n=100 Simulations per Scenario) Estimated using Method-of-Moments

Thus, as a practical matter, in studies powered for ABE (the current international standard, see Chapter 1), inference on variance estimates, or resulting metrics like IBE and PBE, should be approached with caution.

While the previous simulation study was concerned with the scenario where σ_D^2 is in fact non-existent, the more important question is to assess whether estimates of σ_D^2 can be expected to detect subject-by-formulation interaction in cases where it is known to be present. The simulation study depicted in Figure 25 assesses various scenarios based on a range of correlation values, keeping all other factors involved fixed.

The ranges of simulated responses for the estimates of σ_D^2 overlapped despite substantially different scenarios. A correlation of 0.1 ($\rho = 0.1$) indicates very little agreement in subjects' responses between the two formulations, and so we would expect large estimated σ_D^2. However a high correlation of 0.9 would likely not be expected to give rise to large estimated σ_D^2.

This is not the case as 66 out of 100 simulated studies exhibited σ_D^2 in excess of 0.0225 (Figure 25). It was only when the correlation is forced to its most extreme value of $\rho = 1$ indicating perfect agreement between the subjects' responses on the two formulations that the estimated σ_D^2 behaves as expected with 98 of the 100 simulated trials below the cutoff of 0.0225. This would indicate that the estimated value for σ_D^2 in a single bioequivalence study (without the benefit of simulated distributions of studies) may be misleading.

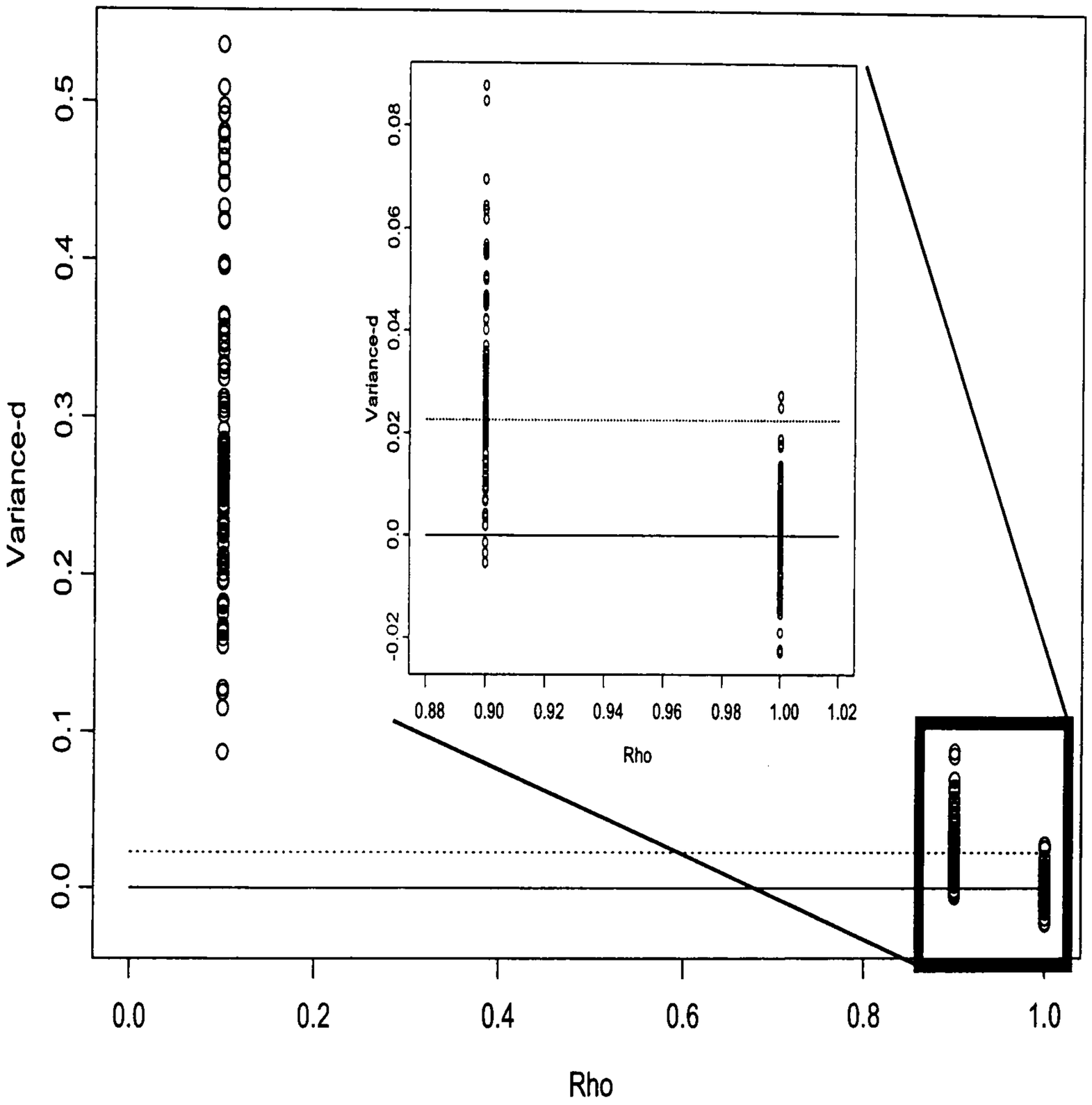


Figure 25: Simulated estimates of σ_D^2 in Replicate Cross-over Studies as a function of ρ when Powered for Average Bioequivalence for a Moderate Variability ($\hat{\sigma}_{WR}^2$=0.04) compound (n=100 Simulations per Scenario, n=54 per simulated study) Estimated using Method-of-Moments

The final simulation study presented in Figure 26 demonstrates that the impact of decreases in correlation (ρ) are dependent upon the magnitude of other variance terms. For low variability drug products, large changes in correlation have negligible effect on the magnitude of $\hat{\sigma}^2_D$, while for products exhibiting higher variation, small changes in correlation can have dramatic effect on the magnitude of $\hat{\sigma}^2_D$. (Figures 25 and 26). This is seemingly counterintuitive since for large variability compounds, lesser agreement should be permissible for the switchability metric.

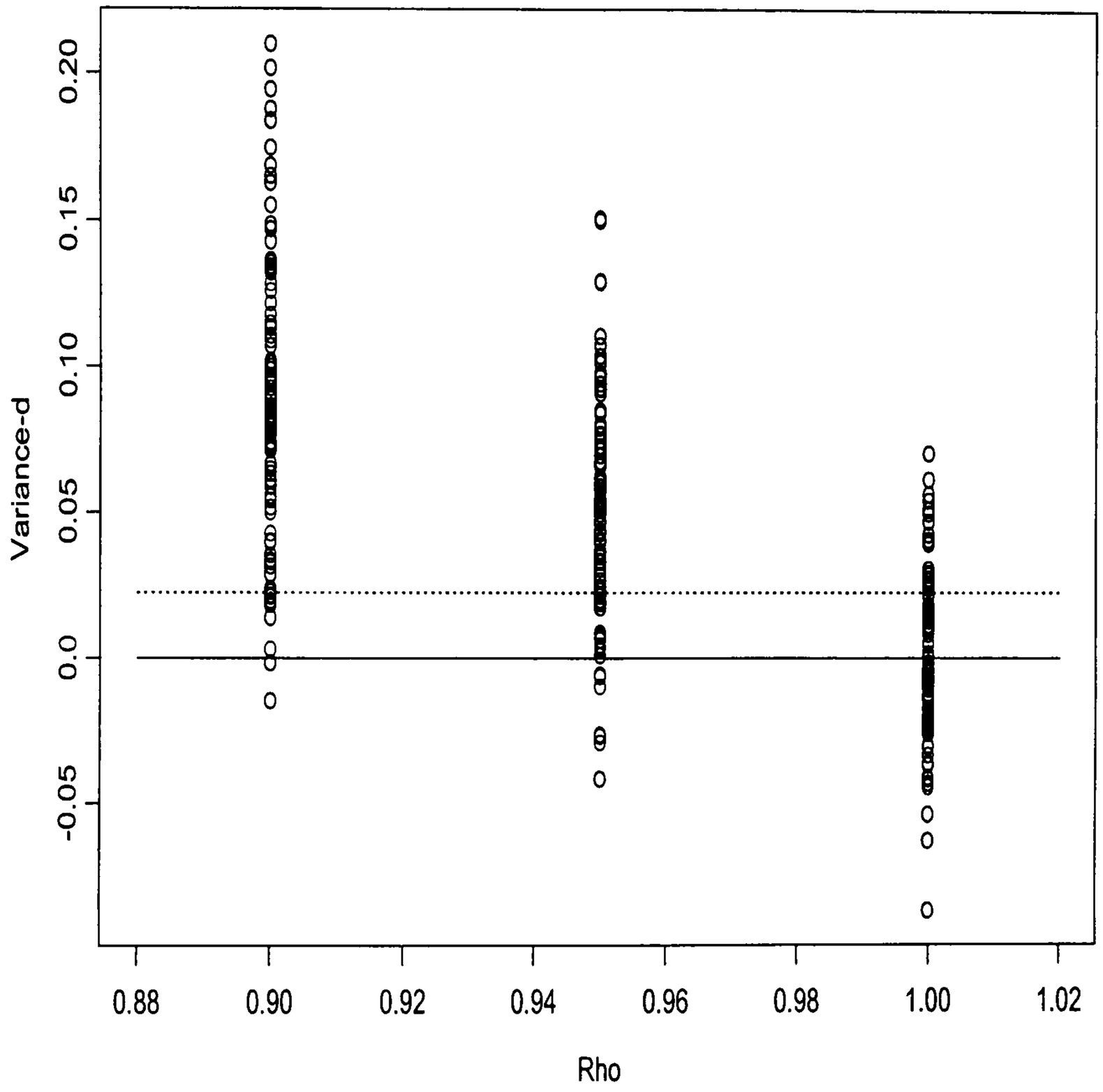


Figure 26: Simulated estimates of σ^2_D in Replicate Cross-over Studies as a function of ρ when Powered for Average Bioequivalence for a High Variability ($\hat{\sigma}^2_{WR}$=0.1225) compound (n=100 Simulations per Scenario, n=54 per simulated study) Estimated using Method-of-Moments

Quantitative definition of a significant subgroup-by-formulation interaction has received little attention in the literature. However, it is known that outliers can occur in bioequivalence studies and such outliers may be indicative of - or mistaken for a subgroup-by-formulation interaction (FDA Guidance, 1992). It should be noted that, in the data sets identified with outliers in our database of 51 studies, retrospective inspection of demographic factors, where available, did not help to explain outlier data (data on file).

We now turn to consideration of the estimates of σ_D^2 from more robust simulation studies tabulated in this thesis.

In complete data sets, method-of-moments and UN REML estimates were mean unbiased regardless of sample size. The CSH REML procedure yielded positively biased estimates when $n = 16$ regardless of whether $\sigma_D^2 = 0$ or $\sigma_D^2 > 0$. When $n = 24 - 80$ we observed that positively biased estimates were found when $\sigma_D^2 = 0$, but that unbiased estimates are derived when $\sigma_D^2 > 0$. RIS REML estimates were positively biased found when $\sigma_D^2 = 0$, but unbiased estimates are derived when $\sigma_D^2 > 0$ for $n = 16 - 80$. In contrast, FA0(2) REML estimates were positively biased when $\sigma_D^2 = 0$, but underestimated σ_D^2 when $\sigma_D^2 > 0$. The mean CSH REML estimates were greater than mean FA0(2) REML estimates across $n = 16 - 80$. Bias in the constrained REML estimates increased as drugs became more highly variable and decreased with increasing sample size (as we would expect in procedures known to be asymptotically unbiased, see Chapter 2-4).

Missing data had a profound effect on estimates of σ_D^2. UN REML estimates remained mean unbiased across $n = 16 - 80$; however, method-of-moments estimates were observed to be positively biased by $0.36 - 0.52$ for $n = 16$, $0.18 - 0.26$ for $n = 24$, $0.11 - 0.16$ for $n = 34$, and $0.04 - 0.06$ for $n = 80$. Bias increased as the drugs became more highly variable and decreased with increasing sample size. CSH REML estimates were positively mean biased to a lesser extent in small ($0.003 - 0.01$ for $n = 16$, $0.01 - 0.04$ for $n = 24$), and the properties of the estimates appeared similar to findings in the complete data sets for larger sample sizes ($n = 34 - 80$). Similarly, mean RIS REML estimates appeared positively biased by $0.002 - 0.09$ for $n = 16$, and the properties of estimates appeared similar to findings in complete data sets for $n = 24 - 80$. FA0(2) REML findings appear similar to those in complete data sets.

For practical purposes, the estimates did not appear in general to differ from a normal

distribution for UN REML estimates sufficiently to cause concern for the UN REML procedure. Normality of the estimates arising from constrained REML procedures (CSH, FA0(2), RIS) in small samples was extremely questionable. This appeared to be due to the poor characterisation of arising estimates, and the constraint placed upon them such that $\sigma_D^2 \geq 0$. Examples are included in Figures 27 and 28 from Simulation Study 1 with $n = 16$ for complete data sets and those with missing data, respectively.

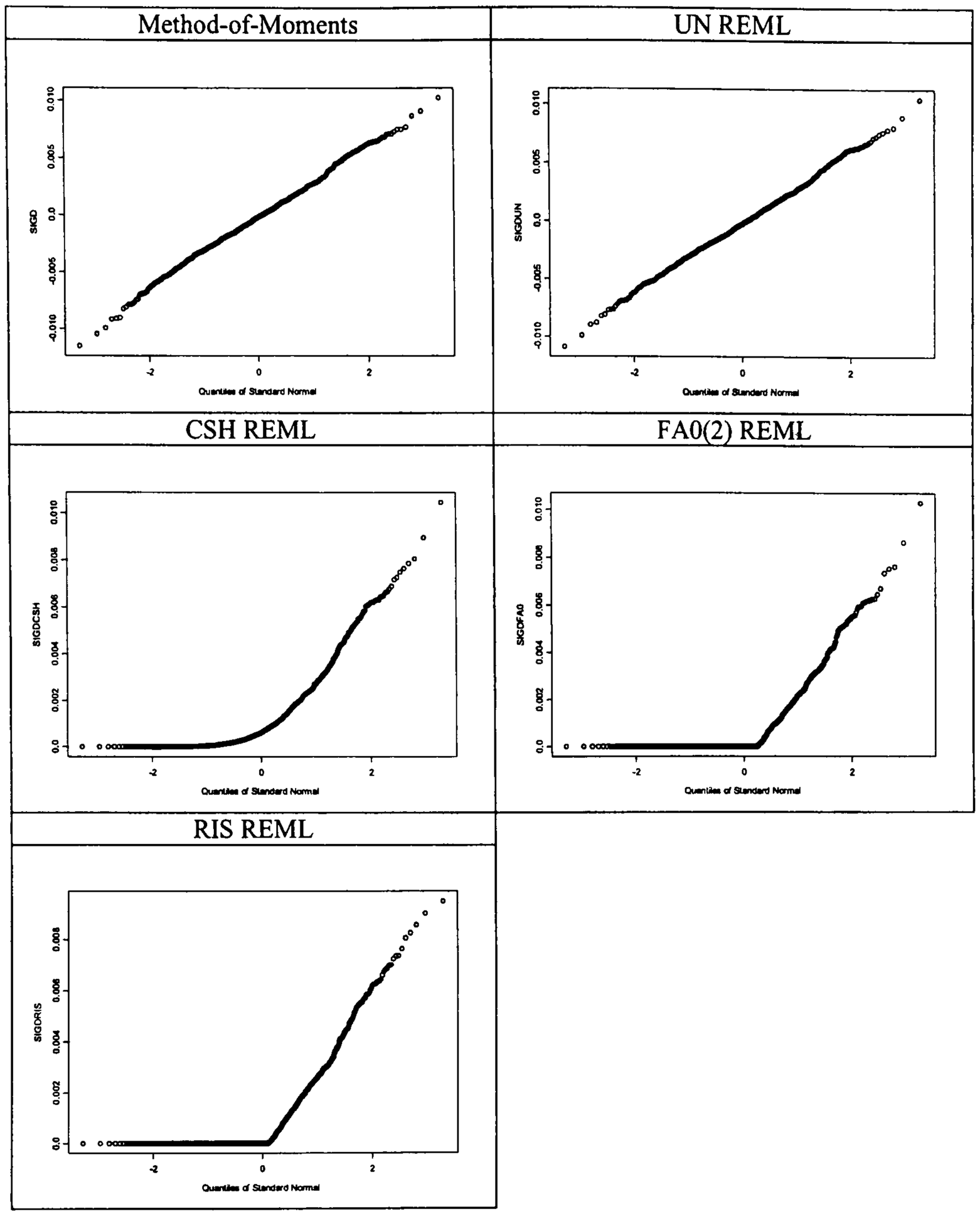


Figure 27: Normal Probability plots for Method-of-Moments and REML $\hat{\sigma}_D^2$ from Simulation Study 1 where $n = 16$ with No Missing data (True $\sigma_D^2 = 0$)

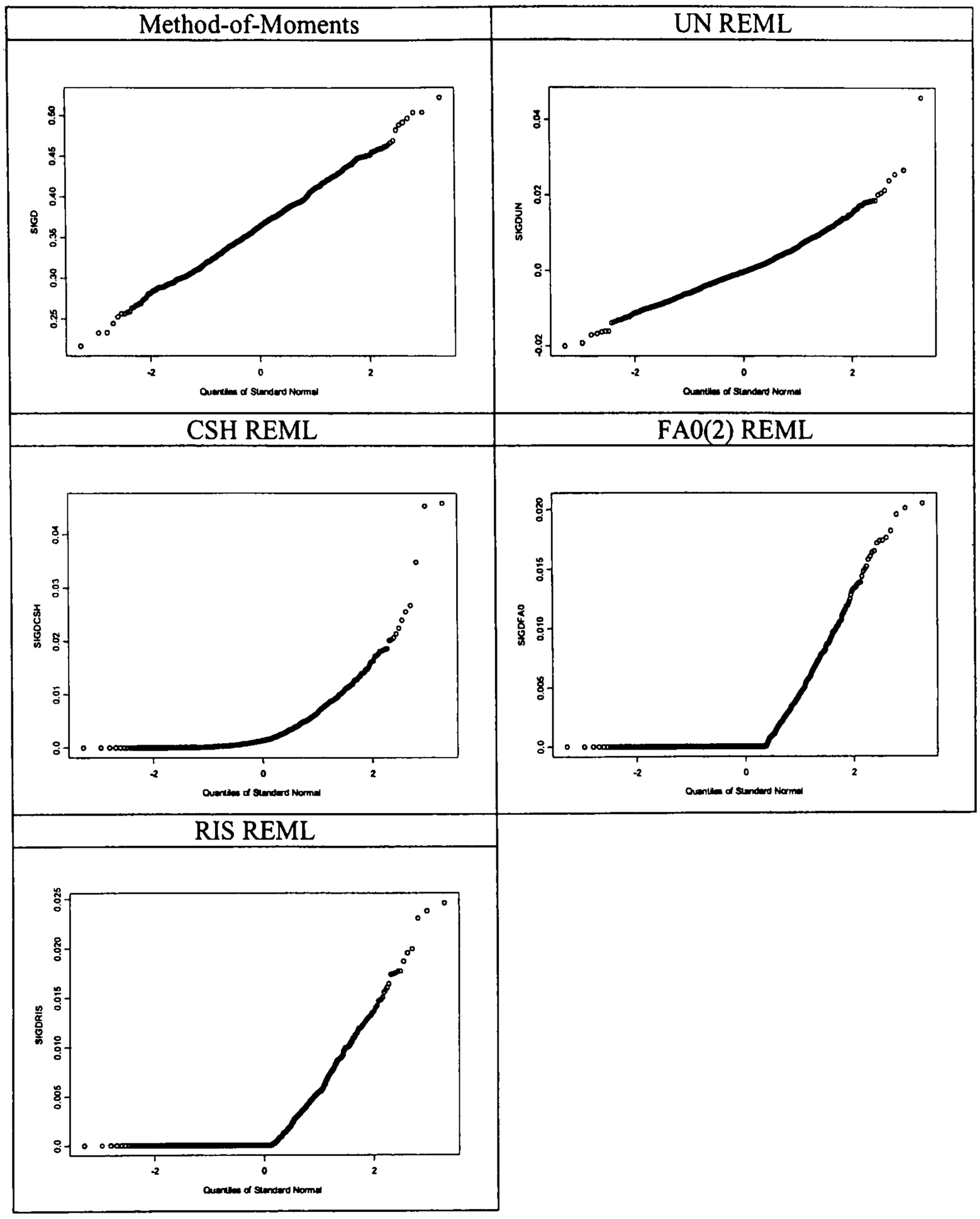


Figure 28: Normal Probability plots for Method-of-Moments and REML $\hat{\sigma}_D^2$ from Simulation Study 1 where $n = 16$ with Substantial Missing data (True $\sigma_D^2 = 0$)

These examples raise several questions. While method-of-moments and UN REML estimates appeared normally distributed, constrained REML procedures CSH, FA0(2), and RIS appear skewed right in distribution (due to the constraints placed on the likelihood). Of interest is whether this pattern continues when $\sigma_D^2 > 0$. To illustrate this, examples are included from Simulation Study 2 with $n = 16$ in Figures 29 and 30 for complete data sets and those with missing data, respectively. This pattern was observed to continue to be present when $\sigma_D^2 > 0$.

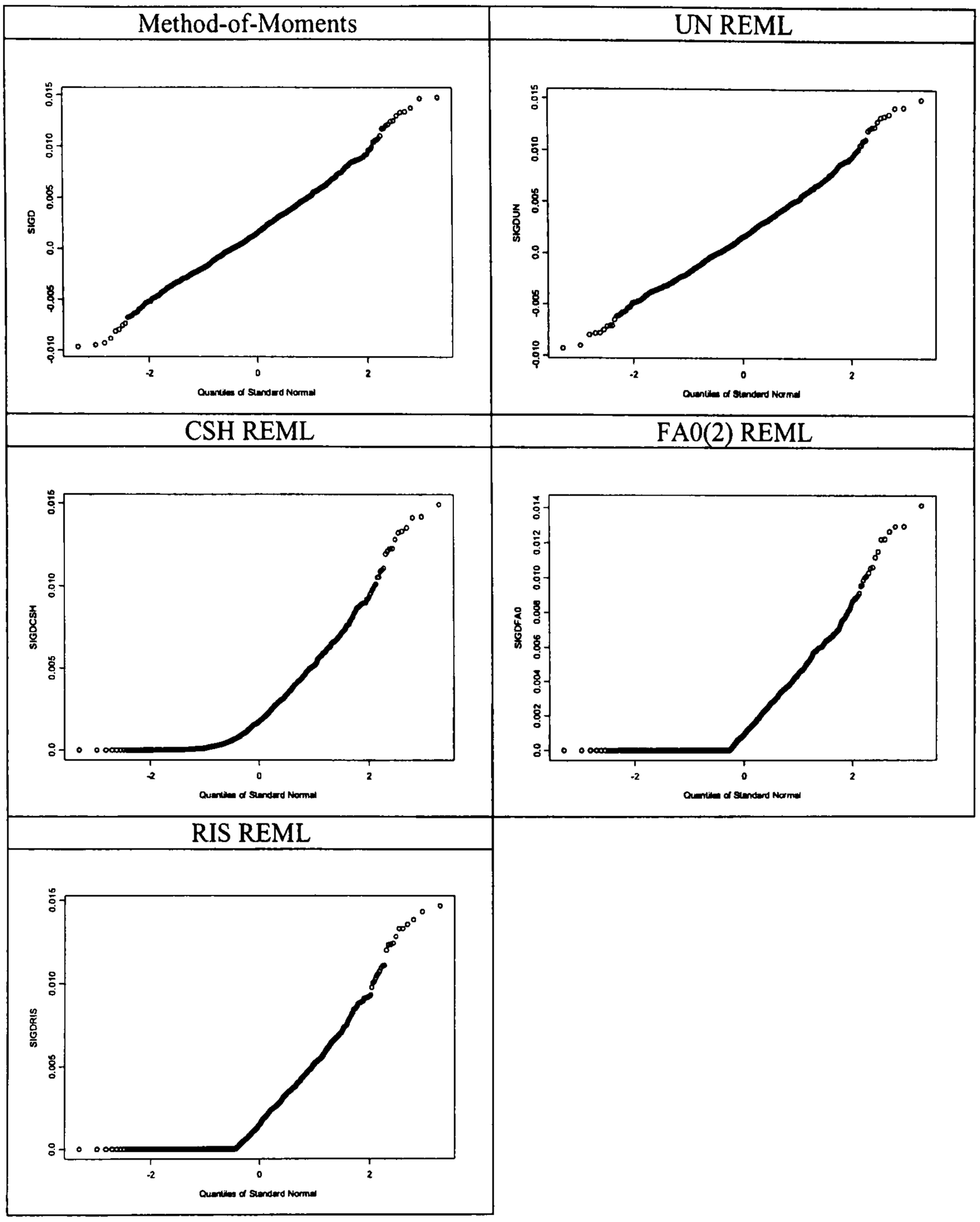


Figure 29: Normal Probability plots for Method-of-Moments and REML $\hat{\sigma}_D^2$ from Simulation Study 2 where $n = 16$ with No Missing data (True $\sigma_D^2 > 0$)

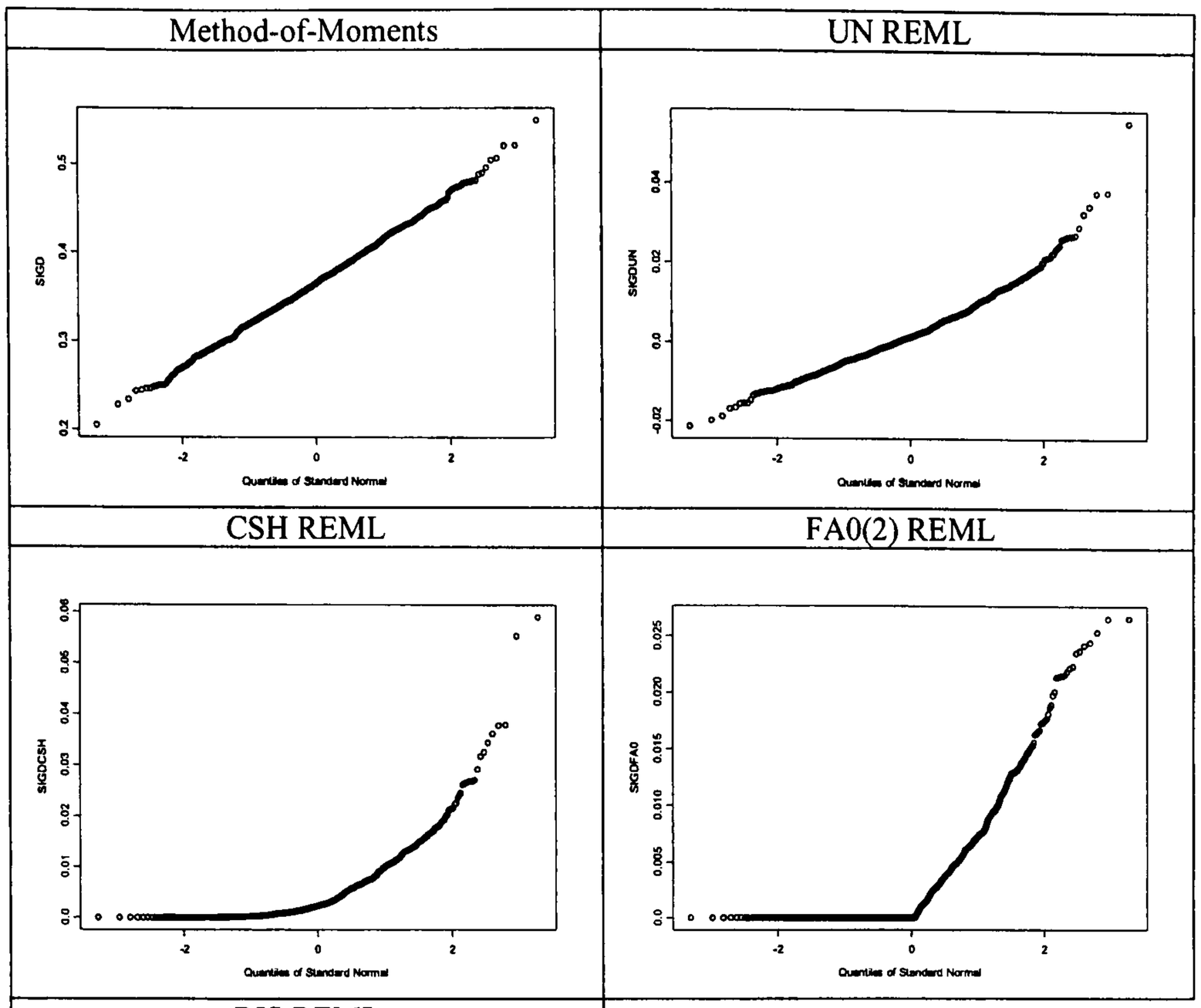


Figure 30: Normal Probability plots for Method-of-Moments and REML $\hat{\sigma}_D^2$ from Simulation Study 2 where $n = 16$ with Substantial Missing data (True $\sigma_D^2 > 0$)

Moreover, it is of interest to see whether this pattern continues with increasing sample size. To illustrate this examples are included in Figures 31 through 34 for Simulation Studies 1 and 2 with $n = 80$. The trends were observed to continue for constrained REML procedures.

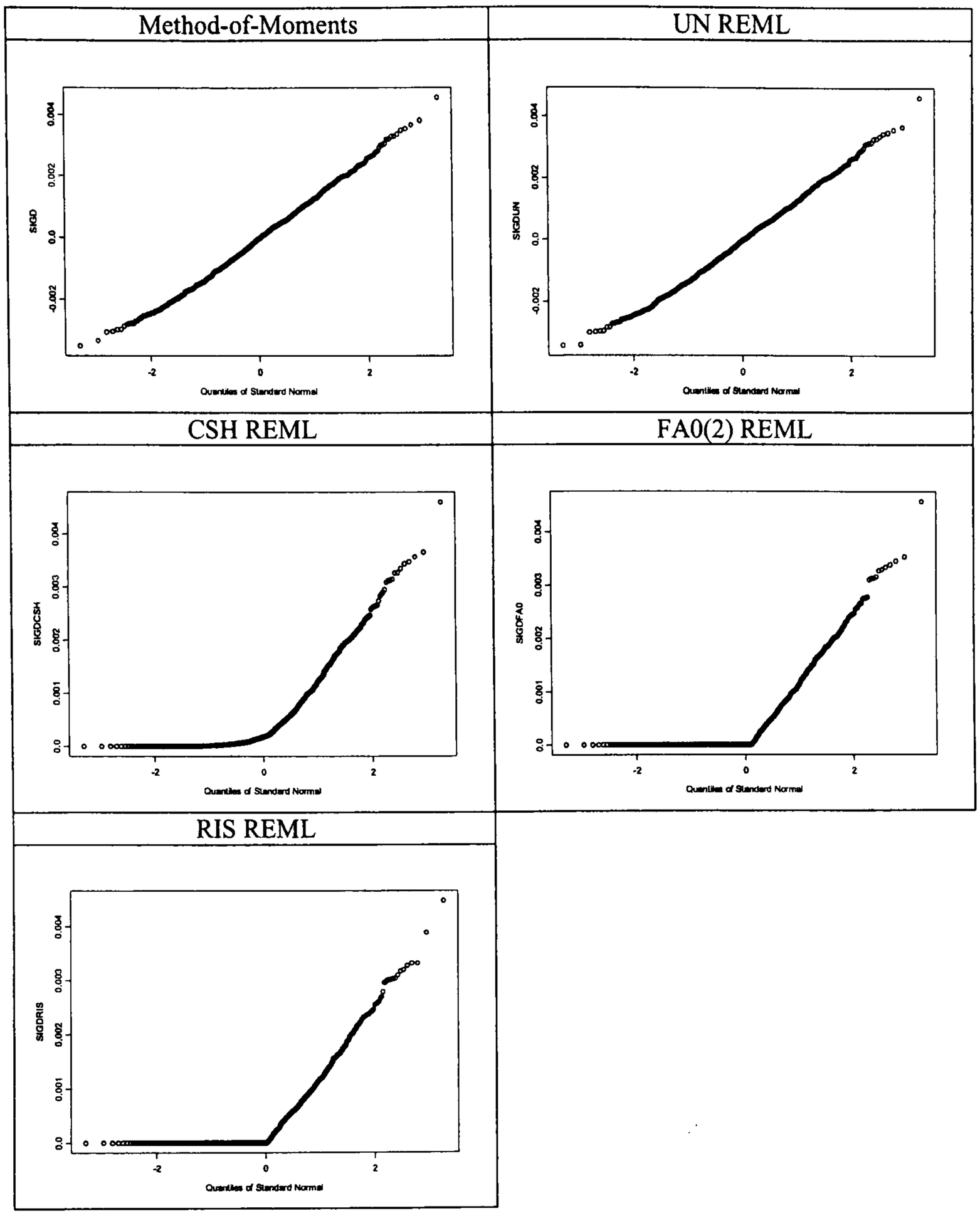


Figure 31: Normal Probability plots for Method-of-Moments and REML $\hat{\sigma}_D^2$ from Simulation Study 1 where $n = 80$ with No Missing data (True $\sigma_D^2 = 0$)

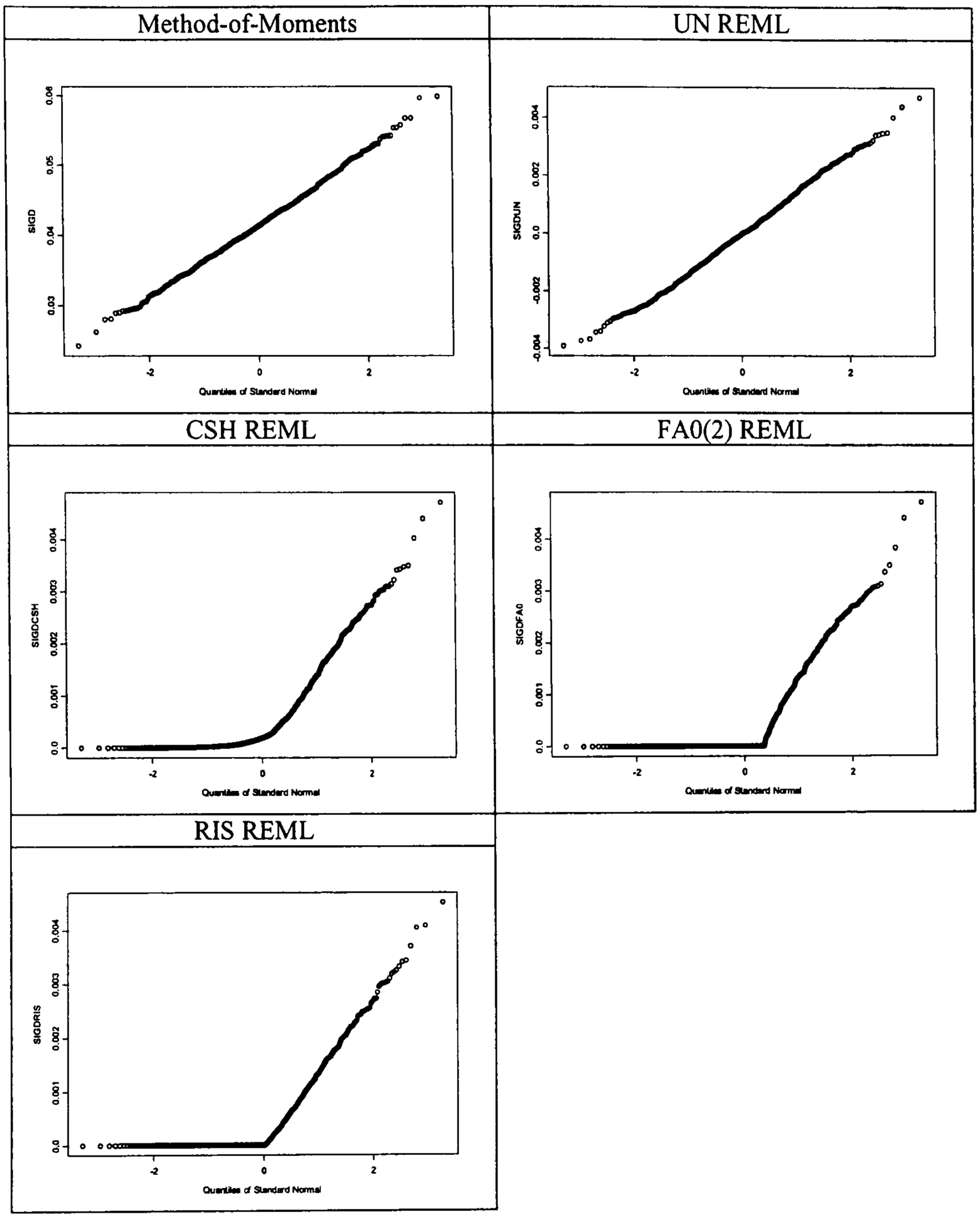


Figure 32: Normal Probability plots for Method-of-Moments and REML $\hat{\sigma}_D^2$ from Simulation Study 1 where $n = 80$ with Substantial Missing data (True $\sigma_D^2 = 0$)

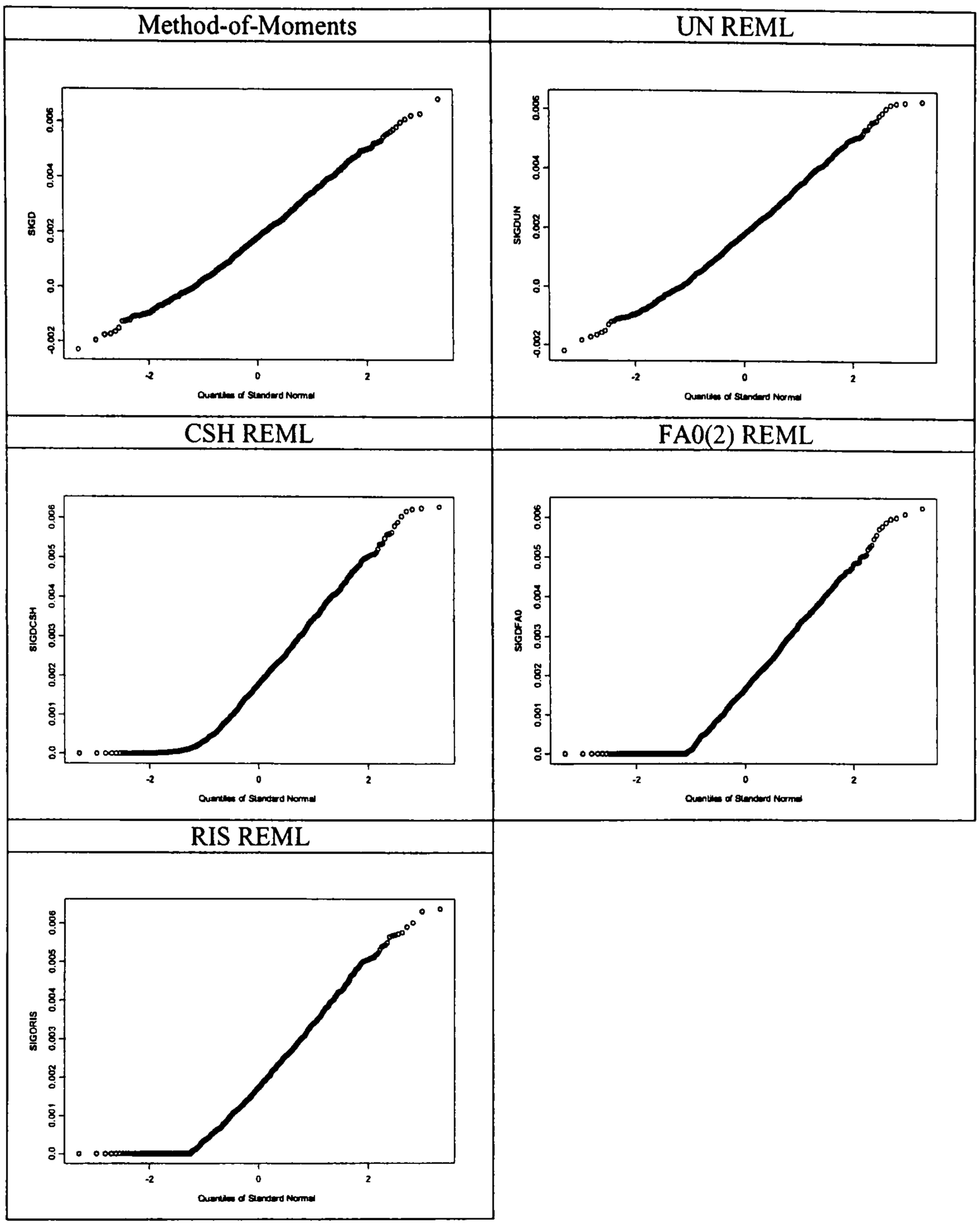


Figure 33: Normal Probability plots for Method-of-Moments and REML $\hat{\sigma}_D^2$ from Simulation Study 2 where $n = 80$ with No Missing data (True $\sigma_D^2 > 0$)

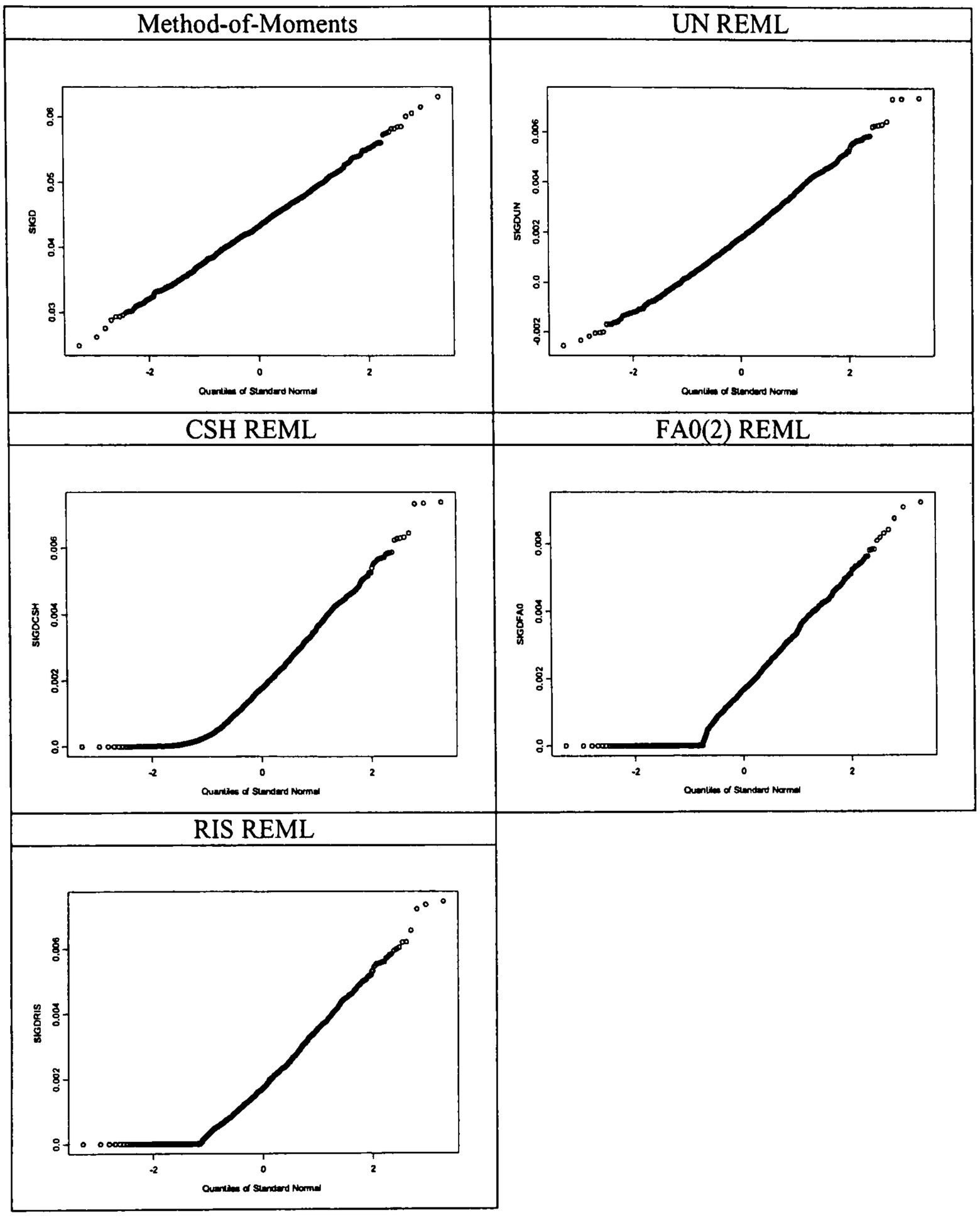


Figure 34: Normal Probability plots for Method-of-Moments and REML $\hat{\sigma}_D^2$ from Simulation Study 2 where $n = 80$ with Substantial Missing data (True $\sigma_D^2 > 0$)

Only one REML procedure (UN) was found to yield unbiased estimates in complete data sets and those with missing data. Method-of-moments, as expected, yielded unbiased estimates in complete data sets, but was positively biased in samples with missing data. Bias in method-of-moments (with missing data) and constrained REML procedures increased as drugs become more highly variable and decreased with increasing sample size. Biased method-of-moments estimates in data sets with missing data were greater than those found in CSH REML which were in turn observed to be slightly greater than those derived using RIS REML. The performance of estimates from FA0(2) REML was questionable. Estimates were positively biased when the true $\sigma_D^2 = 0$ and estimates were negatively biased when $\sigma_D^2 > 0$.

In summary, Subject-by-formulation interaction is strikingly ill-characterized in studies powered to assess ABE and biased estimates are more common than not. Variance associated with this interaction can be generated by a large number of inter-related factors including within-subject variation, sample size, correlation, and between-subject variation. Estimates are biased in the majority of models with bias increasing with variation in the sample space and decreasing with increasing sample size. In general, these observations imply that it will be difficult to separate spurious study results from reality in such designs.

5.3.3 The Pattern of Missing Data and Bias in the Method-of-Moment Estimate of σ_D^2

The pattern of missing data does impact the magnitude of bias in method-of-moments based estimates for σ_D^2. Consider the formula for $\hat{\sigma}_D^2 = M_I - (\frac{\hat{\sigma}_{WT}^2 + \hat{\sigma}_{WR}^2}{2})$ where

$$M_I = \frac{1}{(\sum_{i=1}^{s} n_i) - s} \sum_{i=1}^{s} \sum_{j=1}^{n_i} (I_{ij} - \bar{I}_i)^2$$

and $I_{ij} = \bar{y}_{Tij\bullet} - \bar{y}_{Rij\bullet}$ as previously described. From this formula, it is easy to see that if both observations for the reference or the test product are missing for any given subject, this subject's data will not contribute to the value of I_{ij} nor $\bar{I}_i$ and hence will not bias the estimate of σ_D^2.

However, if one test and one reference is missing for an individual subject, $\hat{\sigma}_D^2$ will be positively biased by an amount proportional to the level of within-subject variation $\frac{a(\frac{\sigma_{WT}^2 + \sigma_{WR}^2}{2})}{(\sum_{i=1}^{s} n_i) - s}$ where a in this expression denotes the number of subjects missing one test and one reference

observation. Given large enough sample size however, we would expect the contribution to bias from this situation to be relatively minor. This finding was confirmed via a simulation sub-study (on file).

Alternatively however, consider the situation where a subjects are missing one test or one reference observation. In this situation, it is easy to see that $\hat{\sigma}_D^2$ derived using method-of-moments will be positively biased by an amount proportional to the between-subject variation and within-subject variation associated with the regimen missing an observation. Between-subject variation can be quite substantial, and this explains our previous findings with regard to the level of bias observed with substantial missing data. Bias would be expected to increase with increasing numbers of subjects with missing data and decrease with increasing study sample size but can be quite substantial even with only 1 missing data point. To illustrate this concept, a simulation sub-study was conducted using Scenario 25. Findings are summarised in Figure 35.

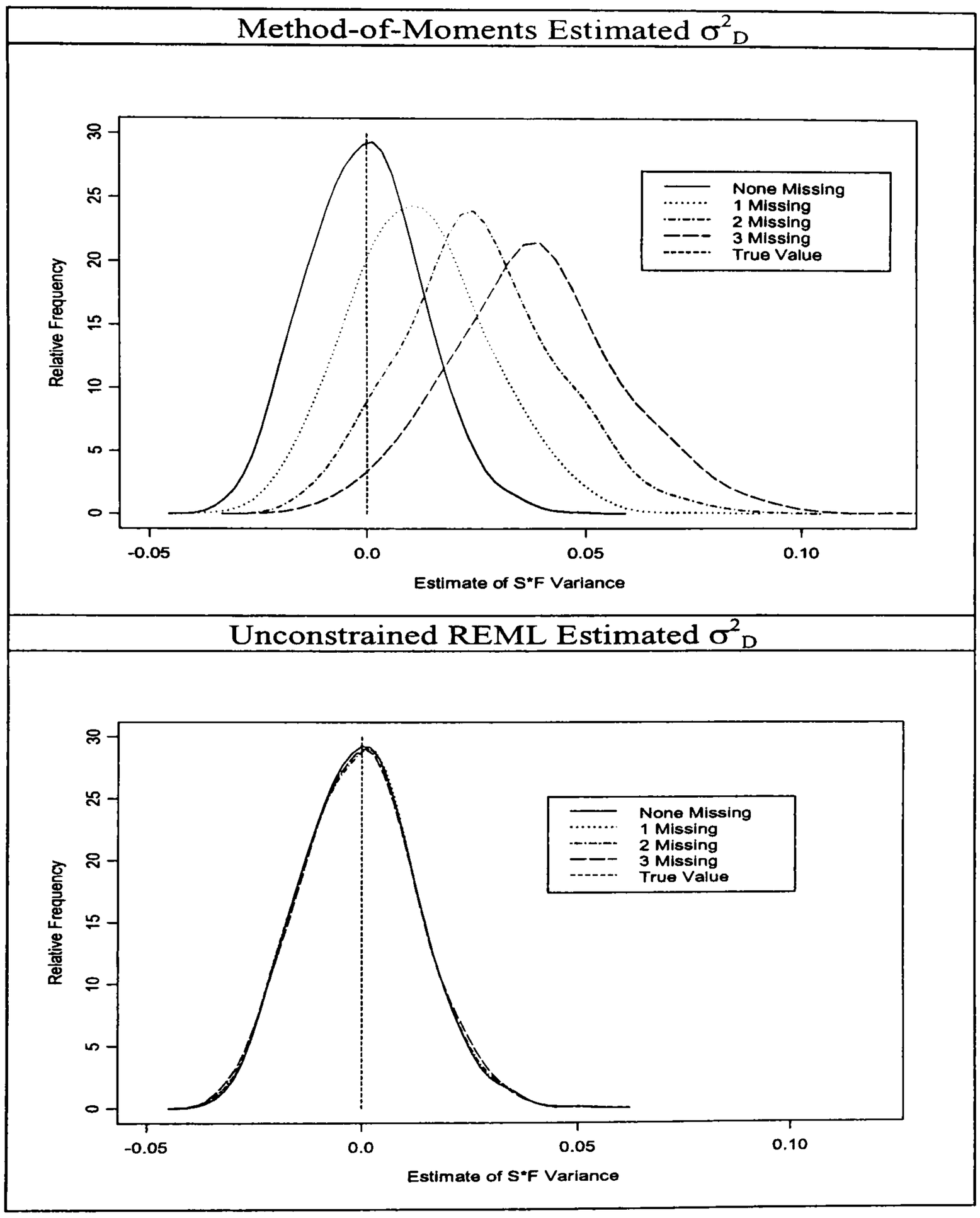


Figure 35: Density Plots for method-of-moments and Unconstrained REML $\hat{\sigma}^2_D$ from Simulation Study 25 where $n = 80$ with $a = (0, 1, 2, 3)$ subjects missing one test product data point

Here we observed that as expected, bias in method-of-moments estimated σ^2_D increased with increasing numbers of subjects with missing data for the test formulation. The variance of the estimates similarly increased with increasing numbers of subjects having one missing data point. However, unconstrained REML estimates continued to be unbiased in distribution and did not appear to be greatly impacted by the presence of missing data in terms of the variation of estimates.

5.3.4 Estimation of σ^2_{WT} and σ^2_{WR}

We begin discussion with assessment of bias in estimates for σ^2_{WT}. Method-of-moments and UN REML were mean unbiased across $n = 16 = 80$ in complete data sets. Bias was small but negative in constrained REML procedures for $n = 16$. For larger sample sizes, constrained REML procedures continued to carry a very slight, but negative, mean bias in estimates of σ^2_{WT} when $\sigma^2_D = 0$. Estimates were mean unbiased when $\sigma^2_D > 0$.

When missing data was introduced, method-of-moments estimates were observed to be unbiased across $n = 16 - 80$; however, a very slight positive bias was observed for UN REML when $n = 16$. However, UN REML estimates were unbiased for $n = 24 - 80$. Constrained REML estimates were observed to be negatively biased when $n = 16$ with mean bias in estimates from CSH and FA0(2) being larger than RIS, and the bias with increasing variation and sample size followed the same pattern as observed for complete data sets.

Estimates for σ^2_{WR} followed the same pattern observed for σ^2_{WT}.

Again, bias increased as drugs became more highly variable and decreased with increasing sample size.

Practical assessment using normal probability plots reveals that discrepancies from normality are relatively minor for UN and method-of-moments estimates as sample size increases beyond $n = 16$ and do not raise practical concerns. See illustration in Figures 36 and 37 below.

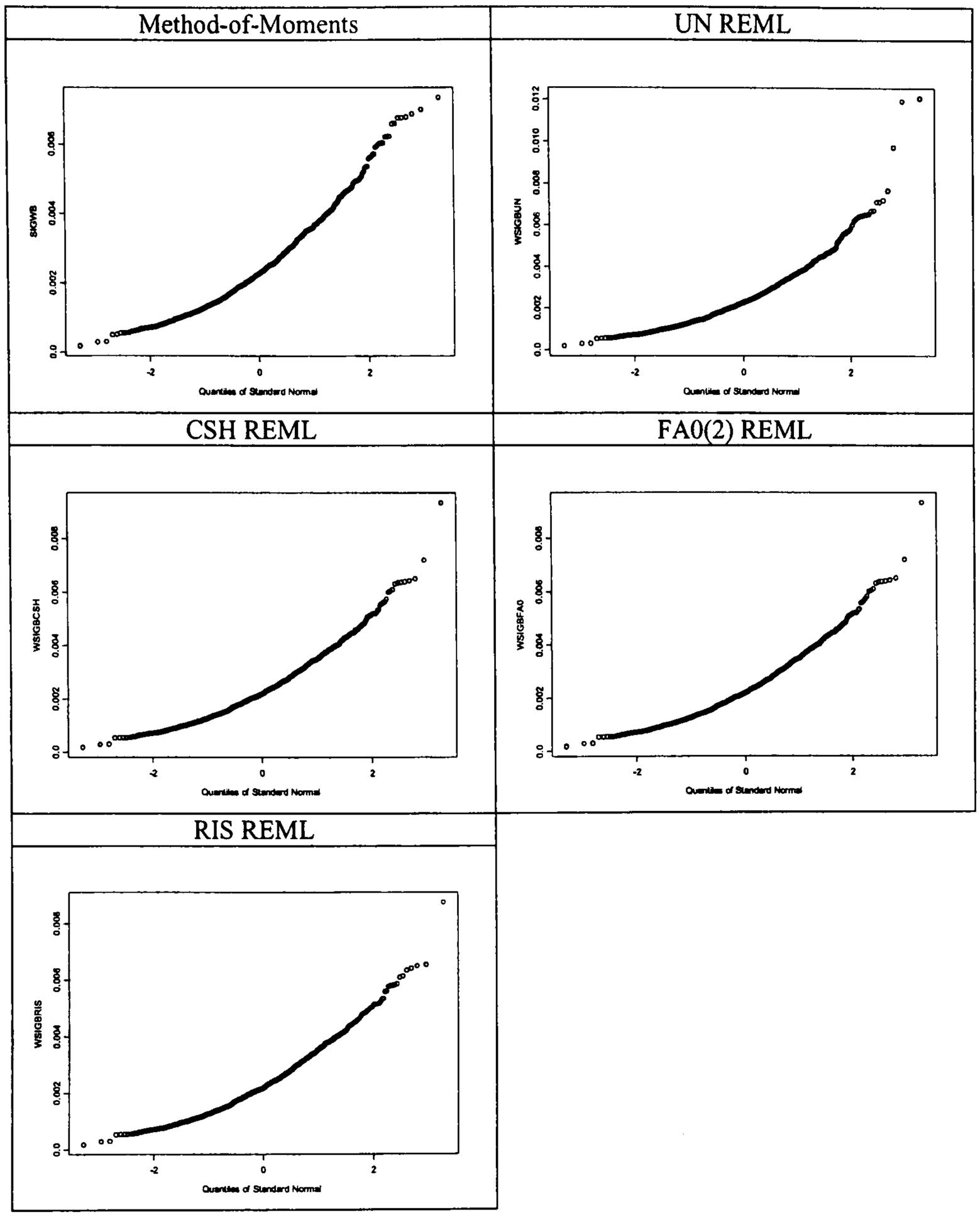


Figure 36: Normal Probability plots for method-of-moments and REML $\hat{\sigma}^2_{WR}$ from Simulation Study 1 where $n = 16$ with Substantial Missing data (True $\sigma^2_{WR} = 0.0025$)

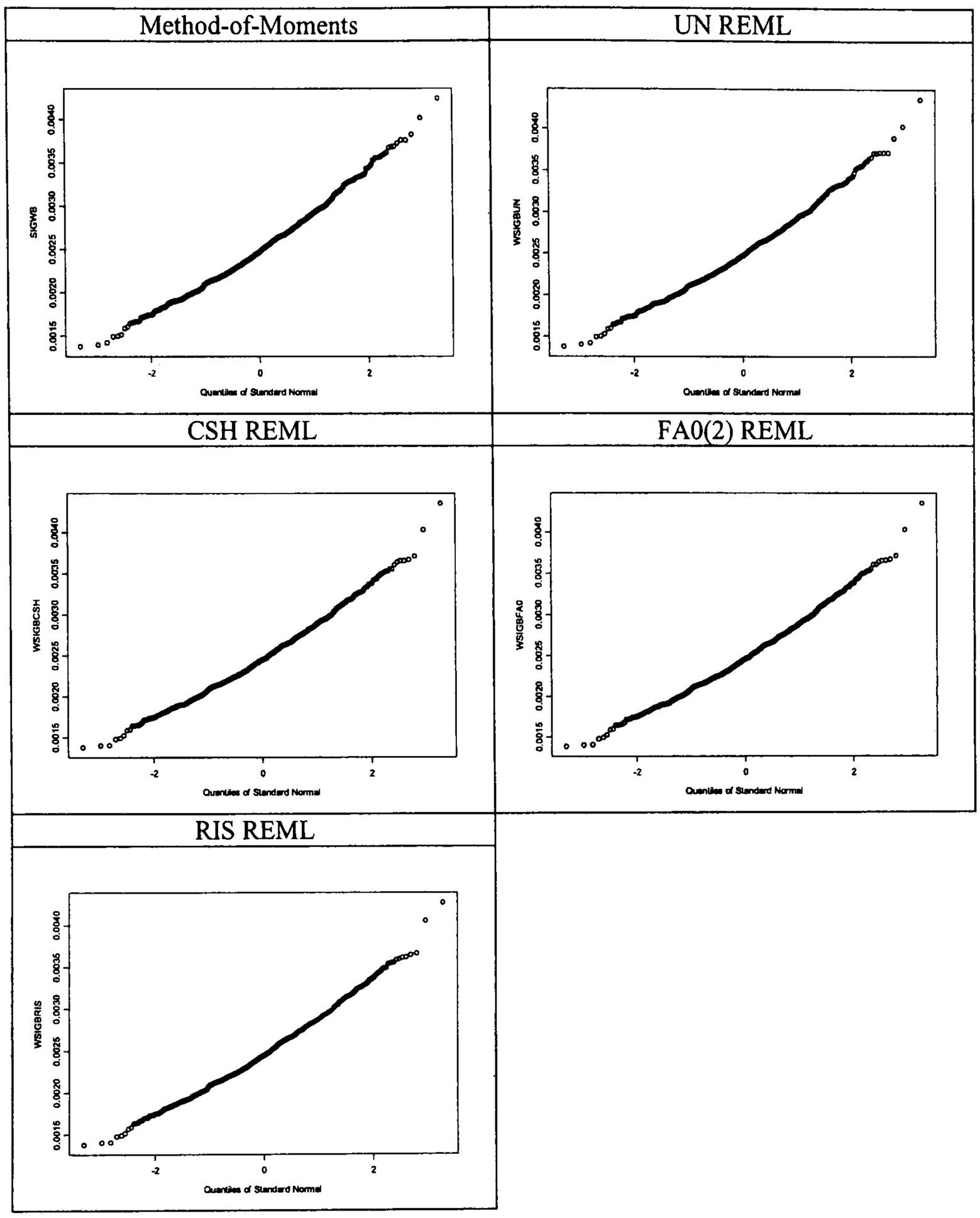


Figure 37: Normal Probability plots for method-of-moments and REML $\hat{\sigma}^2_{WR}$ from Simulation Study 1 where $n = 80$ with Substantial Missing data (True $\sigma^2_{WR} = 0.0025$)

Results were interesting when taken in the context of particular model. Constrained REML estimates are negatively biased for σ^2_{WT} and σ^2_{WR} but are positively biased for σ^2_D. It will be interesting to see whether these findings cancel each other out when assessing ABE. However, it is important to note that the use of the procedures when assessing PBE and IBE is nebulous as bias would be expected to have a non-negligible impact there.

5.3.5 Estimated FDA IBE and PBE Metrics

As expected, in small samples from complete data sets ($n = 16$) positive bias was observed when $\delta = 0$; however, as sample size increases ($n > 24$) this bias becomes smaller. At the maximum samples size studied ($n = 80$), slight negative bias in the estimates for the IBE and PBE metrics were observed on occasion for the method-of-moments and UN REML procedure. This was associated with the mixed scaling procedure for the FDA's metrics and was not unexpected.

As an alternative however, the constrained RIS model was observed to be positively biased for the FDA IBE metric (as expected) when $\sigma^2_D = 0$ and is unbiased only when $\sigma^2_D > 0$. The restricted CSH REML model was studied for the PBE metric; however, as the covariance term associated with σ^2_D does not impact the derivation of the PBE metric, we observed FDA PBE metrics derived using this methods are similarly negatively biased by a small amount for low variability products.

When missing data were introduced, we know from previous findings described in this report that δ^2 and σ^2_D are positively biased when estimated using method-of-moments and are unbiased when using UN REML. These findings carry over to estimation of the FDA IBE metric with positive bias (against sponsors) observed. This bias increased as drugs became more highly variable and was observed to decrease as sample size increased. The unbiased nature of UN REML estimates continued to be prevalent and the biased nature of RIS REML estimates continued when $\sigma^2_D = 0$.

When missing data were introduced, the method-of-moments PBE metric estimates were negatively biased (for sponsors). This bias increased as drugs became more highly variable and decreased with increasing sample size. Exploratory analyses (on file) revealed that this bias was due to positive bias in the total variance estimates for the test formulation co-incident with less degree of positive bias in the reference formulation. In combination, this difference contributed

to the positive bias in the estimated σ_D^2 and negative bias in the function used to compute the estimate for the PBE metric.

Positive bias is generally not of direct concern as it is against sponsors. To compensate, increased sample size would appear to be desirable as this negates such bias and increases precision. This however appears undesirable from the standpoint of decreasing the sample size of these studies due to scaling in use of IBE and PBE compared to ABE (Chapter 1-4).

Negative bias (for sponsors) is of obvious concern to regulators. It too can be negated by increasing sample size, but possibly aids in explaining why it was so easy to demonstrate PBE among the data sets studied in Chapter 4. As an alternative, for nearly unbiased estimation, REML modelling should be used.

5.4 Type I and II Error for Average Bioequivalence from Method-of-Moments and Restricted Maximum Likelihood Models

We begin discussion with regards to Type I error in ABE assessment (Tables 66, 71, 76, 81: simulations 17-54). For $\delta > 0.2231$, all Mom and REML procedures (Tables 66, 71, 76, 81: simulations 37-54) indicated that Type I error rates were 0% in complete data sets and those with substantial missing data. On the boundary point $\delta = 0.2231$, the Type I error rate appeared partially dependent on sample size and the magnitude of σ_D^2.

For complete data sets, method-of-moments and UN REML procedures had Type I error rates less than or equal to 6% when $n = 16$. Type I error rates decreased with increasing sample size, and when $n = 80$, method-of-moments Type I error rate appeared to be less than 5%. However, Type I error rates in the UN REML models were less than or equal to 5% when $n \geq 24$. These findings appeared to be due to the poor characterisation of σ_D^2 and are associated with simulations where the true $\sigma_D^2 = 0$. Type I error rates in excess of 5% occurred when estimates for $\hat{\sigma}_D^2 < 0$, resulting in narrower confidence intervals for approximately half of the simulated data sets.

In contrast, constraints on the estimated parameter space for σ_D^2 constraining it to be greater than or equal to 0 using the CSH, FA0(2), and RIS REML options uniformly constrained the Type I error rate to be less than 5% regardless of sample size.

When missing data were introduced, Type I error rates were near 0 for method-of-moments

based procedures. This was associated with the positive bias in estimates of δ and σ_D^2. As bias in these estimates decreased with increasing sample size, Type I error rates increased. The resulting rates though were less than 5%. Type I error rates for the REML procedures appeared similar to those produced in complete data sets.

We now consider REML modelling and Type II error in ABE testing.

First, it should be noted that in complete data sets REML models are slightly less powerful in some cases that method-of-moments (Tables 66, 71, 76, 81: simulations 1-18). However, in data sets with missing data, the Type II error rate for method-of-moments approaches 100% in small samples (due to the bias in estimates of δ and σ_D^2). Type II error rates in REML procedures appeared roughly the same regardless of constraints placed on the parameter space when $n = 16 - 24$.

When sample size is large $n = 34 - 80$, bias in the estimates of δ and σ_D^2 grew smaller for method-of-moments procedures and Type II error rates began to approach the levels observed for the REML procedures such than when $n = 80$, they were roughly the same.

For complete data sets, when $\sigma_D^2 = 0$, Type I error rates for the method-of-moments procedures are approximately 5-6% in ABE testing. Method-of-moments estimation results in a very biased testing procedure when there is substantial missing data.

The simulated Type I error rate for the UN REML procedure was observed to be approximately 5% in complete and non-complete data sets, and constrained REML procedures result in Type I error rates of less than 5%.

REML modelling was more powerful than method-of-moments modelling in small samples when missing data was present. This finding is likely associated with the bias in method-of-moments estimates of δ and σ_D^2. In larger samples, power for method-of-moments and REML procedures converges and appears similar between procedures. In small samples, method-of-moments modelling should be utilised with caution due to bias in the estimates and the testing procedure. No clear difference in power was evident between constrained and unconstrained REML procedures when testing for ABE.

5.5 Type I Error Rates for Individual and Population Bioequivalence

We now turn to the simulations to assess Type I error in the Hyslop et al. (2000), Asymptotic, and bootstrap procedures.

Simulation findings indicated (see Table 23) that Type I error from the Cornish-Fisher expansion (Hyslop et al., 2000) procedure was approximately 5% for IBE testing; however, the PBE test proposed in the FDA Guidance (2001) appeared biased though conservative with Type I error of approximately 0.5% observed in the simulation studies.

The asymptotic test developed for IBE (Chapter 3) using unconstrained REML estimation also appears biased though conservative except for low variability products (see Simulations 55-59) where rates appeared slightly higher than 5%. As sample size was increased, rates approached 5% for such products. This finding was likely associated with the variation in the estimation of σ_D^2 as described above. When this factor was not accounted for in the asymptotic test for PBE (Chapter 4), Type I error rates appear similar, though slightly less conservative, relative to those found using the Cornish-Fisher expansion. Protection of regulator risk, however, is maintained at less than 5%.

Nonparametric percentile bootstrap procedure findings (using unconstrained REML estimation) do not appear to maintain an acceptable type I error rate with findings indicative of approximately twice the regulatory standard being observed (see Table 23) for low and moderate variability products (Simulations 55-56). Calibration of the confidence intervals to constrain Type I error (see Efron and Tibshirani, Ch. 18, 1993) should be considered if such a procedure is used.

Constrained REML asymptotic testing procedures (see Chapters 3 and 4) appeared slightly more conservative than unconstrained REML asymptotic testing (data on file).

Table 23: IBE and PBE False-Positive Rate (%) from Cornish-Fisher (CF), Large Sample Asymptotic (Asy), and Nonparametric Percentile Bootstrap (NP) Analysis Procedures (1000 runs per Simulation)

Sim	CF	Asy	NP
$n = 16$			
55	4.3	8.6	10.4
56	4.2	3.1	11.1
57	3.3	1.3	3.7
58	3.5	1.2	3.9
59	3.9	1.3	3.8
60	0.6	1.6	6.9
61	0.5	1.1	2.7
62	0.6	1.0	2.2
$n = 24$			
55	4.4	7.7	9.1
56	5.7	3.5	10.5
57	5.1	1.9	4.1
58	5.1	2.2	4.1
59	5.1	2.2	4.1
60	0.6	2.3	5.4
61	0.6	1.9	3.0
62	0.8	1.9	3.1
$n = 34$			
55	4.7	7.3	7.7
56	4.7	4.0	9.5
57	4.1	2.7	4.5
58	4.0	2.8	4.5
59	4.3	2.8	4.7
60	0.3	3.0	4.4
61	0.5	2.8	3.0
62	0.5	2.7	2.9
$n = 80$			
55	3.6	5.6	5.6
56	5.2	3.7	7.8
57	4.7	2.6	4.2
58	5	2.7	4.0
59	5	2.7	4.3
60	0.7	3.9	4.0
61	0.9	3.5	3.5
62	1.1	3.3	3.5

5.6 Type II Error Rates for Individual and Population Bioequivalence

Failure probabilities are tabulated in Tables 90-93 for simulation 1 though 54 in situations involving complete and missing data for sample sizes $n = 16 - 80$ for IBE and in Tables 94-97 for simulation 1 though 54 in situations involving complete and missing data for sample sizes $n = 16 - 80$ for PBE. For the evaluation of Type II error see simulations 3, 5, 9, 11, 15, and 17.

Results for IBE agreed with those found in the retrospective analysis summarised earlier in this chapter, and confirmed that the asymptotic testing procedure developed in Chapter 3 was conservative and less powerful (though this decreased with increasing sample size) relative to the method-of-moments estimation procedure with Cornish-Fisher expansion based inference (Hyslop et al., 2000) in complete data sets. This appeared to be due to the method chosen for estimation of σ_D^2.

This σ_D^2 parameter is estimated in the unconstrained REML model using the between-subject variances of each formulation (σ_{BT}^2 and σ_{BR}^2) and their covariance (σ_{TR}, known to be non-null in cross-over studies) such that $\hat{\sigma}_D^2 = \hat{\sigma}_{BT}^2 + \hat{\sigma}_{BR}^2 - (2\hat{\sigma}_{TR})$. The asymptotic variation (estimated using observed Fisher's information) associated with the covariance ($\hat{\sigma}_{TR}$) is large and when unconstrained, results in very conservative upper bounds for IBE inference. When it is constrained (via constrained REML) to the 'usual' parameter space (i.e. in that $\sigma_D^2 \geq 0$) this asymptotic variance is less, and the upper bound is affected less dramatically.

Alternatively, the properties of method-of-moments estimates for σ_D^2 have been described previously in Chapters 1 and 3; here we remind the reader that in method-of-moments estimation, σ_D^2 is not estimated directly but is estimated as a function of the within-subject variances associated with the estimated difference in formulation means. This measure of variation is less variable than σ_D^2 and is observed to be greater than null on all occasions.

Consider however that in data sets with missing data, the method-of-moments estimates for IBE became quite conservative due to the bias in the estimate $\hat{\sigma}_D^2$, and we found that REML is more powerful for low variability products. This trend decreased with increasing sample size.

If σ_D^2 is not included in a composite metric (i.e. in PBE), we see asymptotic results more as expected relative to the extension of the Hyslop et al. (2000) procedure proposed in the FDA Guidance (2001). Again, this agreed with the findings from our retrospective PBE analysis where we observed that less data sets failed PBE under the asymptotic approach to inference

relative to the Hyslop et al. (2000) procedure. This trend again decreased with increasing sample size and is impacted by the bias introduced with missing data in method-of-moments estimates. However, it is important to note that as sample size increased, it became very easy to demonstrate PBE even when large differences in parameters existed between formulations.

Interestingly, the assumption that $\delta \neq 0$ for the asymptotic IBE and PBE procedures did not appear to impact power dramatically. See Tables 90-97, simulations 1 - 18.

5.7 Transitivity of Individual Bioequivalence and the Potential for 'Drift' in Generic to Generic Switching

We consider the problem of comparability between two generic products (see Chapter 1) approved based on the current IBE criteria (using the Cornish-Fisher approximation technique developed by Hyslop et al., 2000) following the principles of a procedure developed in for ABE by Anderson and Hauck (1996). We wish to answer the question 'How much can average exposure differ between two generic formulations (G_1 and G_2) when they are both declared individual bioequivalent to the same innovator formulation?'

This is to address the practical issue of generic to generic switching in the marketplace pharmacy. When generics enter the market, they are required to be bioequivalent to the innovator formulation, but not to other generic formulations which are also marketed. We know from previous simulations in this chapter that δ can vary quite significantly in the presence of a highly-variable reference product and-or when within-subject variation is decreased on the test formulation.

We will assume that independent individual bioequivalence studies of the generic formulations (labelled formulations G_1 and G_2, respectively) are carried out relative to the innovator formulation. A low variability ($\sigma_{WR} = 0.15$), moderate variability ($\sigma_{WR} = 0.25$), and a high variability innovator product ($\sigma_{WR} = 0.50$) will be considered, and the variation associated with subject-by-formulation interaction will be assumed to be negligible ($\sigma_D^2 = 0$) in both generic studies. Sample sizes will be selected in each study to provide 90% power to demonstrate IBE under these conditions (FDA Guidance, 2001). Sample size will thus be $n = 12$, $n = 32$, and $n = 46$ for the low, moderate, and high variability products, respectively, and for the purposes of study design, we will assume the condition that $\sigma_{WR}^2 = \sigma_{WT}^2$. However, in the simulations

we will allow σ^2_{WT} to vary and examine its potential for impact on the level of allowable δ. It is most likely that as formulation development improves through the life of a drug product intra-subject variation will decrease or remain the same; however, we will also study the potential for increases in the intra-subject test product variance for completeness.

The two generic IBE studies to the reference formulation will be considered independent (as in Anderson and Hauck, 1996). We will assess the probability that G_1 is successfully demonstrated to be IBE to the reference formulation and that G_2 is successfully demonstrated to be IBE to the reference formulation as a function of increasing δ using the Hyslop et al. (2000) procedure in two-sequence, replicate cross-over designs with no missing data. The impact of missing data on IBE testing has been previously discussed and will not be further explored.

The simulation space presented in Table 24 will be utilised.

Table 24: True Values used in Simulation Experiments 63 through 107 (1000 runs per simulation)

Sim	δ_1	δ_2	σ^2_D	σ^2_{WT}	σ_{WR}	N
63	0	0	0	$0.75\sigma^2_{WR}$	0.15	12
64	$\ln 0.95$	$-\ln 0.95$	0	$0.75\sigma^2_{WR}$	0.15	12
65	$\ln 0.90$	$-\ln 0.90$	0	$0.75\sigma^2_{WR}$	0.15	12
66	$\ln 0.80$	$-\ln 0.80$	0	$0.75\sigma^2_{WR}$	0.15	12
67	$\ln 0.60$	$-\ln 0.60$	0	$0.75\sigma^2_{WR}$	0.15	12
68	0	0	0	σ^2_{WR}	0.15	12
69	$\ln 0.95$	$-\ln 0.95$	0	σ^2_{WR}	0.15	12
70	$\ln 0.90$	$-\ln 0.90$	0	σ^2_{WR}	0.15	12
71	$\ln 0.80$	$-\ln 0.80$	0	σ^2_{WR}	0.15	12
72	$\ln 0.60$	$-\ln 0.60$	0	σ^2_{WR}	0.15	12
73	0	0	0	$1.5\sigma^2_{WR}$	0.15	12
74	$\ln 0.95$	$-\ln 0.95$	0	$1.5\sigma^2_{WR}$	0.15	12
75	$\ln 0.90$	$-\ln 0.90$	0	$1.5\sigma^2_{WR}$	0.15	12
76	$\ln 0.80$	$-\ln 0.80$	0	$1.5\sigma^2_{WR}$	0.15	12
77	$\ln 0.60$	$-\ln 0.60$	0	$1.5\sigma^2_{WR}$	0.15	12
78	0	0	0	$0.75\sigma^2_{WR}$	0.25	32
79	$\ln 0.95$	$-\ln 0.95$	0	$0.75\sigma^2_{WR}$	0.25	32
80	$\ln 0.90$	$-\ln 0.90$	0	$0.75\sigma^2_{WR}$	0.25	32
81	$\ln 0.80$	$-\ln 0.80$	0	$0.75\sigma^2_{WR}$	0.25	32
82	$\ln 0.60$	$-\ln 0.60$	0	$0.75\sigma^2_{WR}$	0.25	32
83	0	0	0	σ^2_{WR}	0.25	32
84	$\ln 0.95$	$-\ln 0.95$	0	σ^2_{WR}	0.25	32
85	$\ln 0.90$	$-\ln 0.90$	0	σ^2_{WR}	0.25	32
86	$\ln 0.80$	$-\ln 0.80$	0	σ^2_{WR}	0.25	32
87	$\ln 0.60$	$-\ln 0.60$	0	σ^2_{WR}	0.25	32
88	0	0	0	$1.5\sigma^2_{WR}$	0.25	32
89	$\ln 0.95$	$-\ln 0.95$	0	$1.5\sigma^2_{WR}$	0.25	32
90	$\ln 0.90$	$-\ln 0.90$	0	$1.5\sigma^2_{WR}$	0.25	32
91	$\ln 0.80$	$-\ln 0.80$	0	$1.5\sigma^2_{WR}$	0.25	32
92	$\ln 0.60$	$-\ln 0.60$	0	$1.5\sigma^2_{WR}$	0.25	32
93	0	0	0	$0.75\sigma^2_{WR}$	0.5	46

δ_1: True difference in G_1 to Innovator Formulation
δ_2: True difference in G_2 to Innovator Formulation

Table 24: True Values used in Simulation Experiments 63 through 107 (1000 runs per simulation)

Sim	δ_1	δ_2	σ_D^2	σ_{WT}^2	σ_{WR}	N
94	$\ln 0.95$	$-\ln 0.95$	0	$0.75\sigma_{WR}^2$	0.5	46
95	$\ln 0.90$	$-\ln 0.90$	0	$0.75\sigma_{WR}^2$	0.5	46
96	$\ln 0.80$	$-\ln 0.80$	0	$0.75\sigma_{WR}^2$	0.5	46
97	$\ln 0.60$	$-\ln 0.60$	0	$0.75\sigma_{WR}^2$	0.5	46
98	0	0	0	σ_{WR}^2	0.5	46
99	$\ln 0.95$	$-\ln 0.95$	0	σ_{WR}^2	0.5	46
100	$\ln 0.90$	$-\ln 0.90$	0	σ_{WR}^2	0.5	46
101	$\ln 0.80$	$-\ln 0.80$	0	σ_{WR}^2	0.5	46
102	$\ln 0.60$	$-\ln 0.60$	0	σ_{WR}^2	0.5	46
103	0	0	0	$1.5\sigma_{WR}^2$	0.5	46
104	$\ln 0.95$	$-\ln 0.95$	0	$1.5\sigma_{WR}^2$	0.5	46
105	$\ln 0.90$	$-\ln 0.90$	0	$1.5\sigma_{WR}^2$	0.5	46
106	$\ln 0.80$	$-\ln 0.80$	0	$1.5\sigma_{WR}^2$	0.5	46
107	$\ln 0.60$	$-\ln 0.60$	0	$1.5\sigma_{WR}^2$	0.5	46

δ_1: True difference in G_1 to Innovator Formulation
δ_2: True difference in G_2 to Innovator Formulation

The output of these simulated experiments will be the percentage of successful paired (but independent) studies for given levels of $\delta_2 - \delta_1$ and σ^2_{WT}. This will allow us to determine when one should take caution when switching from a generic formulation to another generic formulation at the pharmacy. A similar problem was address under the ABE criteria by Anderson and Hauck (1996). Given the conclusions of Chapters 4 and the simulations carried out in this Chapter with regard to the PBE criteria, we recommend PBE not be used to allow market access as proposed and did not study its potential impact in this setting.

The findings of these simulations are summarised in Figures 38, 39, and 40 for low, moderate, and high variability products, respectively.

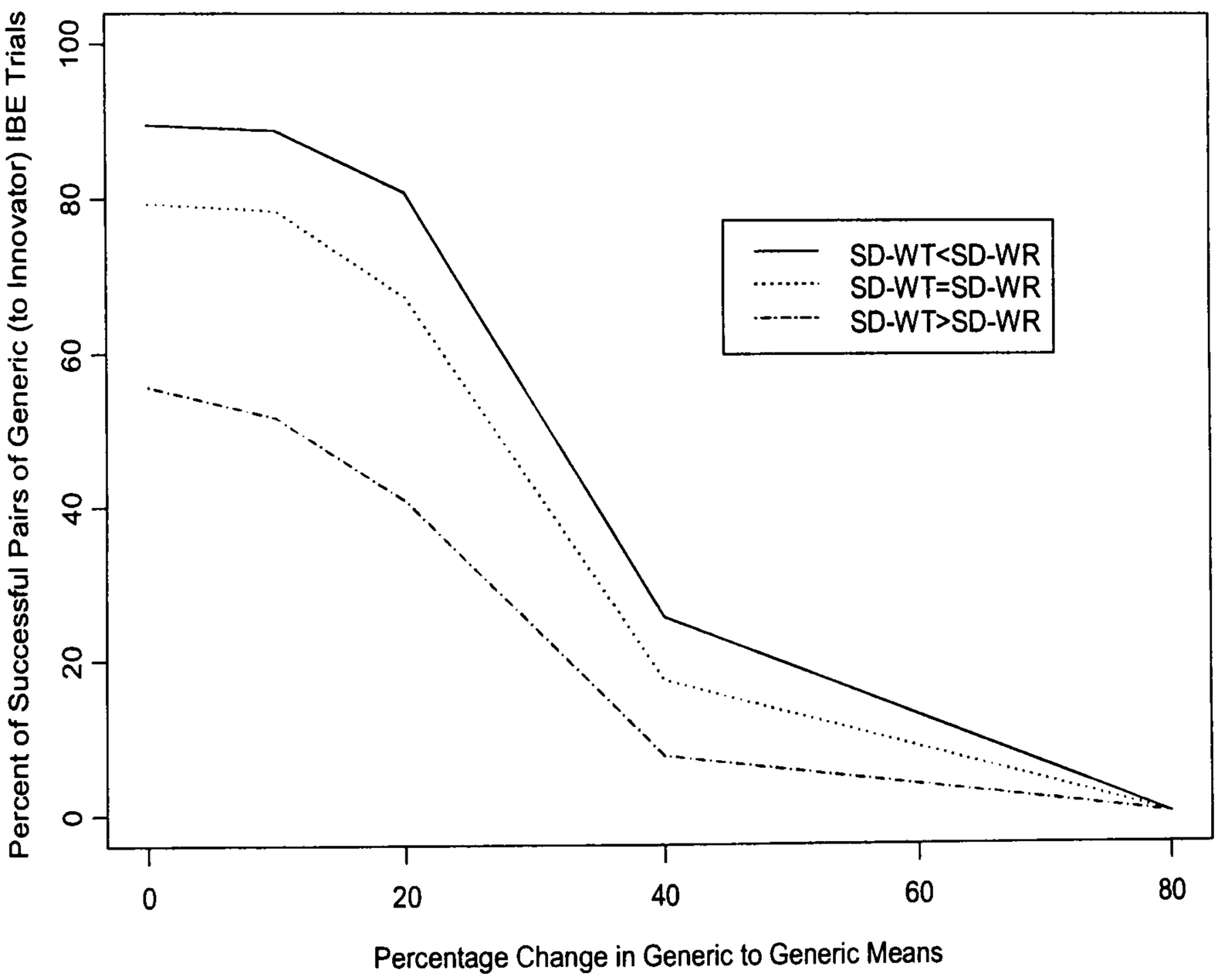


Figure 38: Percentage of Pairs of Independent Generic to Innovator IBE trials Demonstrating IBE versus the True Difference in Generic to Generic Means for a low variability innovator product $\sigma_{WR} = 0.15$ in 1000 simulations ($n = 12$ for each simulated study)

For the low variability innovator product in Figure 38, it is observed that for test (Generic) products with decreased or equal within-subject variability mean exposure may vary between generic products up to 10% with some frequency (around 80% of the simulated studies). Changes in mean exposure up to 20% in generic products (the traditional threshold of clinical concern) will occur 60-80% of the time when switching from generic to generic produces under the proposed method from FDA. Changes in mean exposure up to 30% may occur with some frequency (i.e. 50%). Changes in mean exposure in excess of 30% are possible but unlikely. Note that if the test product has increased within-subject variability, mean exposure between generic products was only observed to differ by up to 10% in around 50% of the studies.

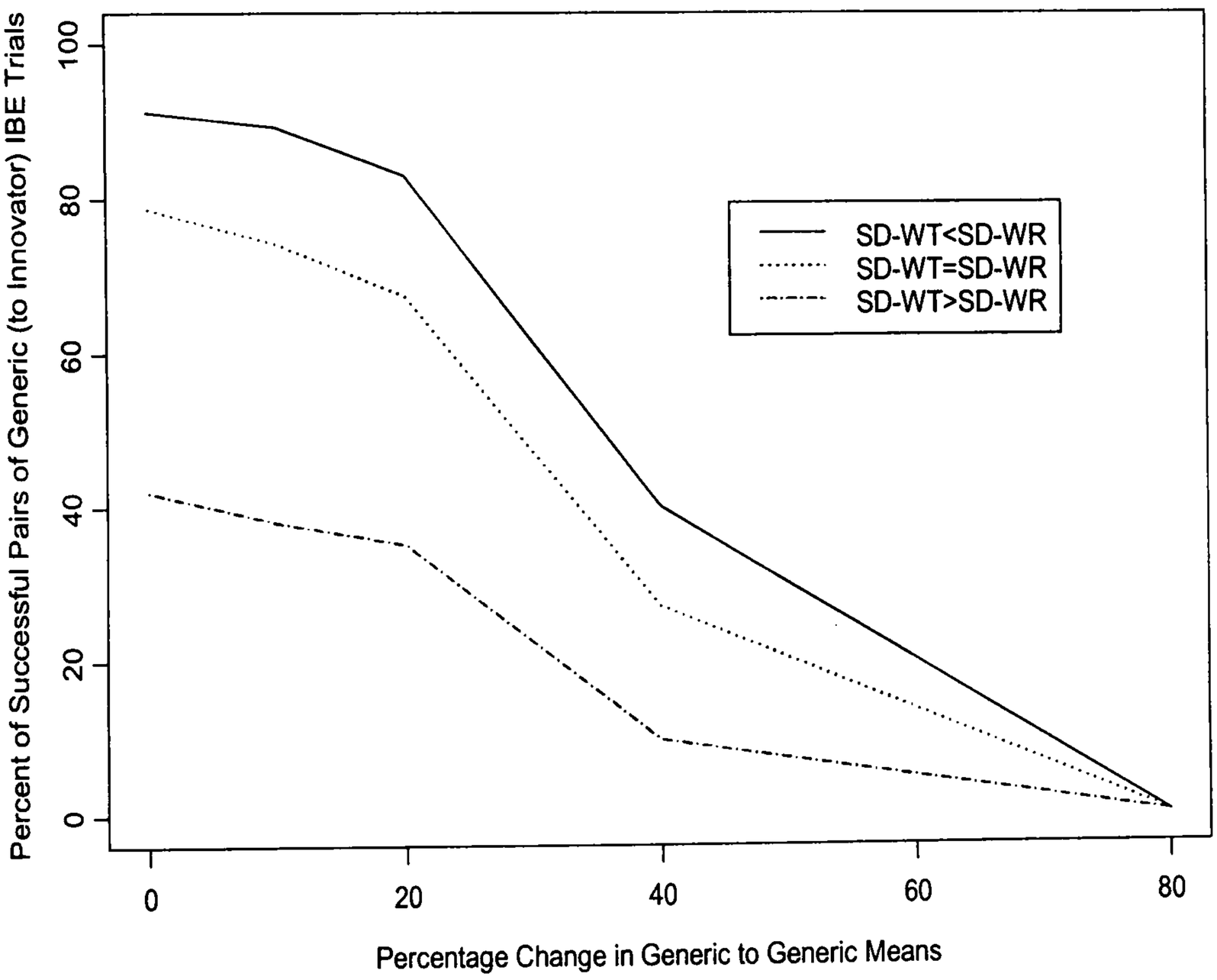


Figure 39: Percentage of Pairs of Independent Generic to Innovator IBE trials Demonstrating IBE versus the True Difference in Generic to Generic Means for a moderate variability innovator product $\sigma_{WR} = 0.25$ in 1000 simulations ($n = 32$ for each simulated study)

For products with moderate within-subject variation in the innovator product, it is observed that for test (Generic) products with decreased or equal within-subject variability mean exposure may vary between generic products up to 20% with some frequency (around 80% of the simulated studies). Changes in mean exposure up to 35-40% may occur with some frequency (i.e. 50%). Changes in mean exposure in excess of 30% are possible but unlikely. Note that if the test product has increased within-subject variability, mean exposure between generic products was only observed to differ by up to 10% in less than 50% of the studies.

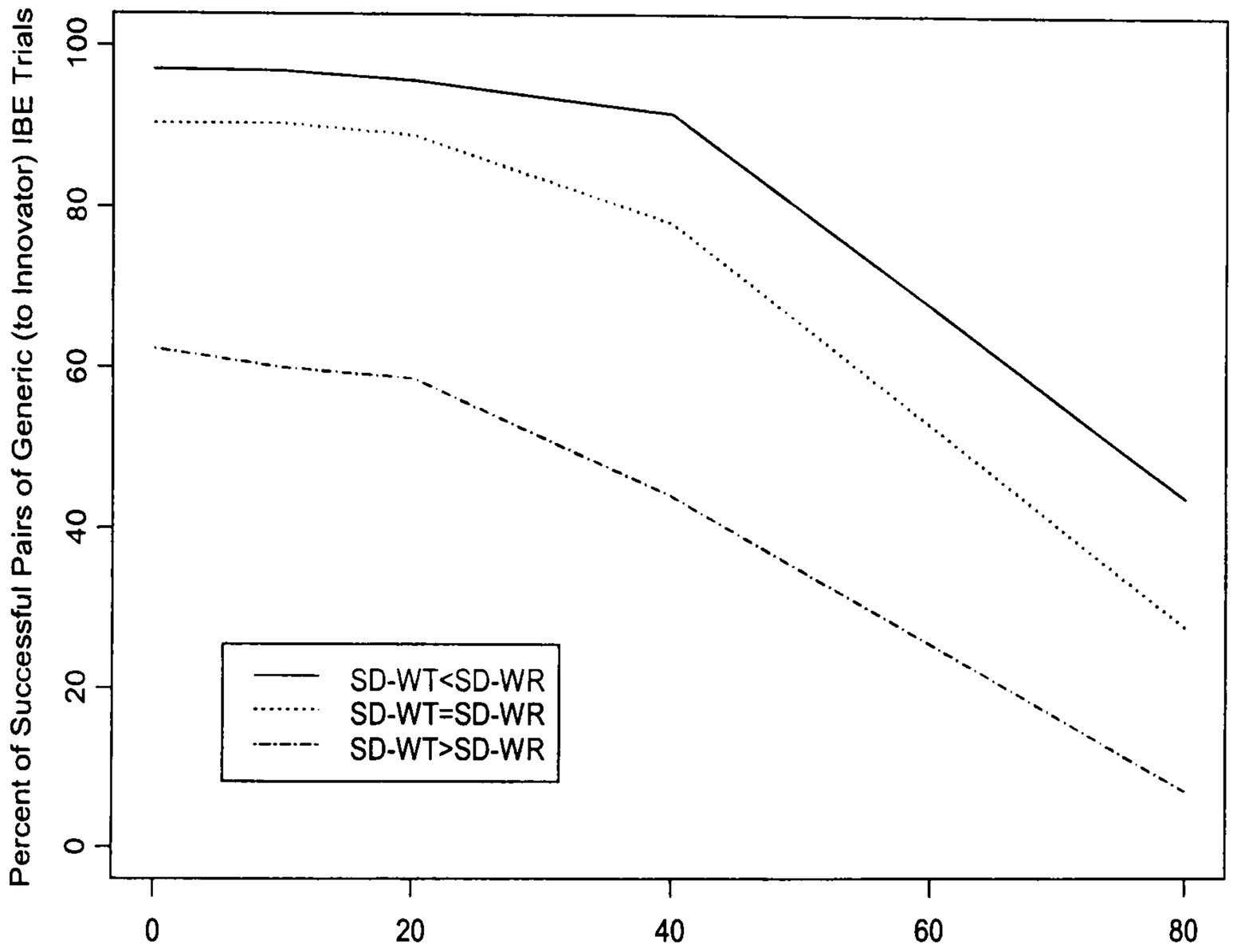


Figure 40: Percentage of Pairs of Independent Generic to Innovator IBE trials Demonstrating IBE versus the True Difference in Generic to Generic Means for an high variability innovator product $\sigma_{WR} = 0.5$ in 1000 simulations ($n = 46$ for each simulated study)

For products with high within-subject variation in the innovator product, it is observed that for test (Generic) products with decreased or equal within-subject variability mean exposure may vary between generic products up to 40% with some frequency (around 80% of the simulated studies). Changes in mean exposure up to 60% may occur with some frequency (i.e. 50%). Changes in mean exposure in excess of 60% are possible with some frequency. Note that if the test product has increased within-subject variability, mean exposure between generic products was only observed to differ by up to 20% in approximately 50% of the studies.

5.8 Discussion

Average bioequivalence will likely continue to serve as the international regulatory standard as its properties are well understood. IBE and PBE have yet to be shown to be valid for the same purposes, and evidence of therapeutic failure under the ABE approach has yet to surface. In small samples, especially those with missing data, method-of-moments modelling for ABE should be utilised with caution due to bias in the estimates and the testing procedure.

The constrained REML procedure recommended by FDA Guidance (2001) using Sattertwaite (1941) degrees of freedom for ABE testing in replicate designs results in biased estimates for variance components on occasion; however, it uniformly constrains the rate of Type I error (of more immediate concern to Regulators and Consumers) to be less than 5% in ABE testing. Thus the FDA Guidance (2001) tacitly acknowledges that 'While all models are wrong, some are useful.' If Kenward and Roger's (1997) approach to estimation of the degrees of freedom is used, the degrees of freedom are the same as those found using Satterthwaite's procedure; however, the variance estimate for the confidence interval for δ is inflated using the approach of Harville and Jeske (1992) to account for shrinkage in the fixed and random effects from the mixed modelling procedure to provide a confidence interval with at least 90% coverage probability. This results in confidence bounds for δ which are slightly larger than those found when Satterthwaite's approach is used and leads to slightly more conservative Type I error rates for ABE than those observed using the Satterthwaite option in $SAS^{©}$. As these rates already protect public health risk (as shown in this Chapter using simulation), Kenward and Roger's (1997) procedure protects regulatory risk and may also be applied to test for average bioequivalence.

The bias observed in method-of-moment estimates of δ when missing data are present is

unlikely to be of direct concern in ABE testing. It is common practice (see FDA Guidance, 2001) to plan for deviations in δ of up to 0.05 when planning a trial. Moreover, the FDA Guidance (2001) recommends the use of a REML procedure for ABE assessment, which we know to be asymptotically unbiased and have found via simulation is in general unbiased for δ in small samples.

The potential method-of-moments bias in δ induced by missing data is more of immediate concern in PBE and IBE assessment under the recommended (FDA Guidance, 2001) approach to analysis. While the slight positive bias in small samples induced by the 'plug-in' procedures recommended in the FDA Guidance (2001) is not of sufficient magnitude to be of concern (see Chapters 3 and 4), when biased estimates for δ or σ_D^2 are 'plugged' into the metric, bias is of concern and may impact inference.

Variance estimates are of less concern in ABE testing, but in alternative criteria where estimates are important to interpretation (i.e. for IBE and PBE) method-of-moments estimates should be viewed cautiously. Method-of-moment estimation, as expected, yields unbiased estimates in complete data sets, but results in positively biased $\hat{\sigma}_D^2$ in some samples with missing data. Bias in method-of-moment $\hat{\sigma}_D^2$ (with certain patterns of missing data) and constrained REML procedures increases as drugs become more highly variable and decreases with increasing sample size. Biased method-of-moment estimates in data sets with missing data exhibit a greater degree of bias than those found in CSH REML, and the estimates from an alternative constrained (FA0(2)) REML are similarly questionable. Only the unconstrained REML procedure (Type='UN') was found to yield unbiased estimates for σ_D^2, σ_{WT}^2, and σ_{WR}^2 in complete data sets and those with missing data.

Research continues to indicate that subject-by-formulation interaction variance (σ_D^2) is strikingly ill-characterized in studies powered to assess ABE, and biased method-of-moment estimates are common in the presence of missing data. Variance associated with this interaction can also be generated by a large number of inter-related factors including within-subject variation, sample size, correlation, and between-subject variation (Hauck et al., 2000; Zariffa and Patterson, 2001). Simulation studies indicate that method-of-moments estimates may be biased with the magnitude of bias increasing as drugs become more highly variable and decreasing with increasing sample size. In general, these observations continue to imply that it will be difficult to

separate spurious study results for σ_D^2 from reality in such designs.

In situations where this is of concern, asymptotic based IBE and PBE inference appears similar to that of the Cornish-Fisher expansion recommended by FDA and provides a practical, though conservative alternative to the FDA procedure. There is precedent for the use of asymptotic tests in the study of pharmacokinetics (Machado et al., 1999), but such techniques should be viewed with caution and studied thoroughly using simulation before their use becomes widespread. Such procedures should be considered when the pattern of missing data is sufficient to cause concern with the potential for bias in method-of-moments estimation. Harville and Jeske's (1992) and Kenward and Roger's (1997) approach to estimation in this setting could be extended to provide a more precise conservative upper bound in this setting for the asymptotic testing procedure developed in Chapters 3 and 4; however, as simulation has shown in this report that the asymptotic testing procedure is in general conservative, such extension was not considered here.

Turning to the statistical testing procedures, the Cornish-Fisher expansion procedure for PBE (FDA Guidance, 2001) is flawed in theory due to it not taking account of covariance among the estimates; however, when the asymptotic procedure is applied to take account of such in data sets with missing data (and hence non-null covariances), little to no practical difference in inference is observed. Based on the simulation studies reported in this paper, both asymptotic and Cornish-Fisher tests appear to protect the Type I error rate. It is unlikely that accounting for these covariances is of pivotal importance to PBE inference though the validity of PBE itself as an adequate protection for public health is questionable as proposed.

It is very easy to demonstrate PBE for most products regardless of the presence of large changes in mean exposure between formulations, and the FDA's Guidance (2001) permits a potential problem to surface in public health due to generic-to-generic pharmacy switches if approval is granted based on IBE as proposed.

In conclusion, the findings of this chapter suggest that the Cornish-Fisher expansion (Hyslop et al., 2000) will adequately serve for IBE and PBE testing except in the presence of missing data where method-of-moments estimates become biased. In situations where missing data and the resulting bias in estimates are of great concern, an asymptotic testing procedure using REML (though conservative) may be used to assess inference.

While valid statistical tests for PBE have been developed under the proposed FDA standards, this procedure quite easily allows for market access with very large changes in mean exposure for highly variable drug products. The potential for threats to public health generated by generic-to-generic switching should not be underestimated if IBE is used to allow market access. We recommend that the FDA reconsider the use of the IBE and PBE procedures for market access and not allow their use without major modification to ensure patient safety and efficacy.

6 Population Pharmacokinetics and Bridging

The findings of this chapter were presented at the Societe Francaise de Statistique (Patterson et al., 2001f), at the Drug Information Association meeting (Patterson et al., 2001i). Aspects of the findings were published in a series of a GlaxoSmithKline reports (Altman et al., 2000 and Patterson et al., 2002d) and in a GlaxoSmithKline technical report (Patterson and Jones, 2002i).

6.1 Background on Bridging

Population bioequivalence (see Chapters 1, 4, and 5) is defined as the assessment of bioequivalence for formulations used in pivotal clinical trials relative to the to be marketed formulation and falls under the definition of prescribability. This approach is said to answer the question, 'Can a patient (naive to drug) be exposed to either the clinical trials formulation of the to-be-marketed formulation with equal assurance that the efficacy and safety profile established in the clinical trials will 'hold true' for them also?' (FDA Guidance, 2000b)

In such a situation, it is assumed that the population exposed to drug in the pivotal clinical trial is a subset of that to be exposed when the drug is marketed. 'Population' bioequivalence is thus somewhat of a misnomer in that the formulation is being changed in the study of interest (and not the population, the change for which happens outside the study). In any event, however, the practice of referring to this type of bioequivalence as 'population' is well established (FDA Guidance, 2000b-2001) and will be abbreviated in this thesis as PBE, where P refers to 'Population'.

In the assessment of PBE, it is assumed that the clinical trials population is a subset of the full population. It is held (FDA Guidance, 1997, 1999a-b, 2000b) that cross-over designs for

PBE in a normal healthy volunteer population are sufficient to isolate any differences between formulations. Thus, cross-over or replicate cross-over designs (Jones and Kenward, 1989; Senn, 1993; Senn, 2002) are considered appropriate for use in these studies, and the properties of PBE in such studies were developed in great detail in the previous chapters. It is generally the case that no quantitative evaluation is made regarding the appropriateness of this assumption that the clinical trials population is a true and representative subset of the market population. Such evaluations are qualitative are left to the purview of the regulatory agency from whom authorisation to market is sought. For some agencies, it may be of interest to ensure that the assessment of equivalence in formulations is not confounded with any uncontrolled change in population.

As a first step then, consider an assessment of 'true' bioequivalence between differing populations when the same formulation is given to differing populations (or ethnic groups). In this context, it is informative to explore the International Conference on Harmonisation (ICH-E5, 1998) 'Guidance on Ethnic Factors in the Acceptability of Foreign Clinical Data'. Work by the author on these topics may be found in Patterson et al. (2002i) and work by the author was included in a points to consider document in the implementation of this ICH-E5 guidance (Altman et al., 2000). These materials are included for completeness, and these topics will be expanded upon in this thesis, concentrating upon application to pharmacokinetic studies.

The purpose of this chapter is to develop model-based techniques that may be used in situations involving small samples and when the inclusion of covariate information may be important to assess whether pharmacokinetics are equivalent between ethnic groups. We first consider a clinical development strategy for 'bridging' data from one population into another and then concentrate in the subsequent section on issues particular to experimentation involving pharmacokinetic endpoints. Next we develop methods appropriate for the comparison of populations derived from the FDA PBE metric and an alternative procedure developed based on the Kullback-Leibler Distance (Dragalin and Fedorov, 1999a) and illustrate their use using a data set from a recent submission and simulation. The approaches which are developed will be illustrated using an existing data set and explored using simulation.

6.2 Clinical Drug Development Planning for Inter-Regional Bridging

It is common for any new drug to be studied primarily in a particular region, such as North America and Europe. The initial requests for registration are made to the regulatory authorities within these original regions, and in order to accelerate global market access and accelerate patient access to the latest improvements in health care, it is desirable to be able to utilize these data in registering the product in other regions such as Pan-Asia. These new regions, however, must be assured that that regional and ethnic differences do not impact the product's safety, efficacy, dose, and dose regimen. The stated purpose of ICH-E5 (1998) is "to facilitate the registration of medicines among ICH regions by recommending a framework for evaluating the impact of ethnic factors upon a medicine's effect, i.e., its efficacy and safety at a particular dosage and dose regimen."

As described in ICH-E5, the first of two primary requirements for a submission package is that the data requirements for registration in the new region be met - i.e. that clinical trial methodology, record-keeping, protocol compliance and drug accountability, and informed patient consent must be acceptable in the new region (ICH-E5, 1998). The minimal data package, consisting of either data from the original region and/or data from the new region, should include an adequate characterization of the pharmacokinetics (PK), pharmacodynamics (PD), dose response, efficacy and safety of the drug (see Chapter 1 for more details). At least PK (Naito, 1998a), and preferably PD and dose response, should also be characterized in a population that is relevant to the new region (ICH-E5, 1998) but not necessarily resident in the new region (Naito, 1998b).

The second requirement (the domain of 'Bridging' bioequivalence, henceforth referred to as BBE) is the demonstration of the ability to extrapolate findings from any data from the original region to the population of the new region. It is easier to extrapolate from one region to another if the new medication is "ethnically insensitive," i.e., unlikely to behave differently in different populations. Ethnic sensitivity can be categorized into two components, intrinsic (genetic) and extrinsic (environmental), either or both of which may impact bioavailability and hence the appropriate dose and response relationship. These are described in greater detail in the following Figure 41.

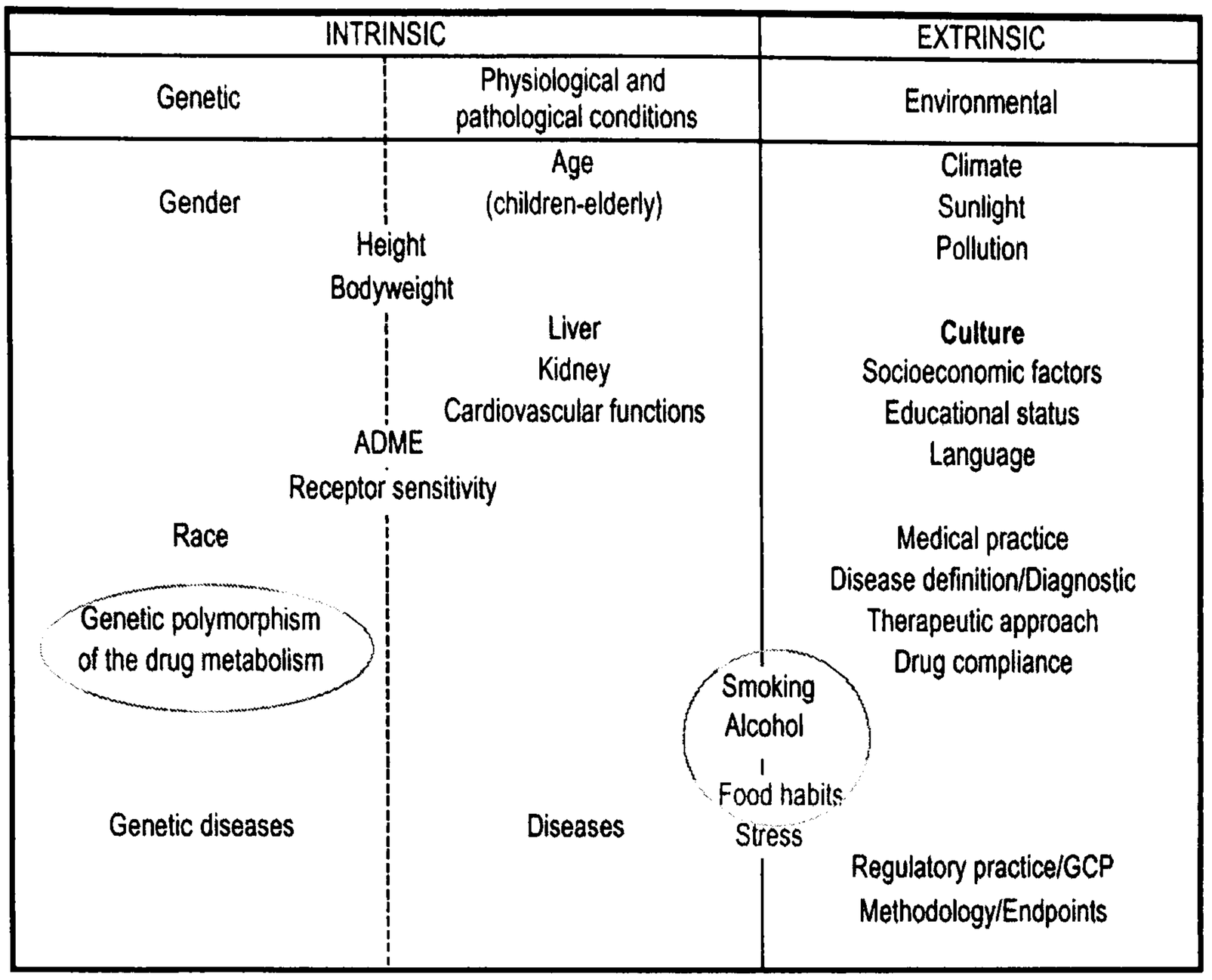


Figure 41: Intrinsic and Extrinsic Population Factors Impacting the Dose-PK-PD Response Relationship (ICH-E5, 1998)

A 'bridging' study, as its name implies, is designed to allow one to bridge from the original region's (foreign) data in the original population to the new region. It is a (ICH-E5, 1998):

> "supplemental study performed in the new region to provide pharmacodynamic or clinical data on efficacy, safety, dosage, and dose regimen in the new region that will allow extrapolation of the foreign clinical data to the new region... Extrinsic ethnic factors are factors associated with the environment and culture in which a person resides. Extrinsic factors tend to be less genetically and more culturally and behaviourally determined. Examples of extrinsic factors include the social and cultural aspects of a region such as medical practice, diet, use of tobacco, use of alcohol, .. compliance with prescribed medication and practices in clinical trial design and conduct."

The degree of ethnic sensitivity will determine whether a study is necessary and the design of such a study (e.g. PK only, PK/PD only, in what population, etc.). ICH-E5 (1998) describes several characteristics of drug products which would make such a product 'ethnically insensitive'. These are:

> "1. Linear pharmacokinetics
> 2. A flat response curve for both efficacy and safety in the range of the recommended dosage and dose regimen (this may mean the medicine is well tolerated)
> 3. A wide therapeutic dose range (again an indicator of good tolerability)
> 4. Minimal metabolism or metabolism distributed among multiple pathways
> 5. High Bioavailability, thus less susceptibility to dietary absorption effects
> 6. Low potential for protein binding
> 7. Little potential for drug-drug, drug-diet, and drug-disease interactions
> 8. Nonsystemic mode of action
> 9. Little potential for inappropriate use"

It is very rare for a drug to meet all nine conditions which would make it ethnically insensitive and result in only minimal data requirements to enter new regions and markets (e.g. such as Pan-Asia). However, early phase pharmacokinetic studies of the drug product will hopefully help assess which of the nine conditions may not be met and allow for an informed decision to be taken with regard to how to bridge inter-ethnic data as part of clinical studies in patients in global drug development.

We now turn to points to consider in the incorporation of an ICH-E5 bridging strategy in a clinical development program. The strategy spans across studies in normal volunteers and clinical studies in patients in the original filing regions. Differences between populations in intrinsic and/or extrinsic factors (see above) may impact measures of central tendency or variation in pharmacokinetics by altering a drug product's absorption, distribution, metabolism, or elimination in different populations. These changes may or may not result in concentrations

at the site of action sufficient to elucidate a clinically meaningful change in safety or efficacy response between populations.

Prior to the first-in-man study in clinic, in-vitro and animal pre-clinical experimentation should establish a range of safe doses for study in man. Doses are then selected for introduction into clinical studies in Phase I in humans (see Reigner and Blesch, 2001). Clinical development of a drug product, with the exception of only the most toxic products targeted for the treatment of cancer, then initiates with the study of the drug product in normal healthy male volunteers in Europe or the USA. These studies are typically small, well-controlled, data-intensive, dose escalating, and placebo-controlled. In this stage of development, the primary objective of a clinical study is to determine a safe range of doses and dosing regimens (e.g. once-a-day or twice-a-day) for later dosing in studies involving patients with the disease state under study. Dose and dosing regimen are examined with respect to their impact on the PK of the drug product, and additionally, should biomarker or surrogate markers (Biomarker Definitions Working Group, 2001) be present to characterize the pharmacodynamic activity of the drug in normal healthy volunteers, these data are characterized relative to dose or PK levels. By the end of Phase I, dose finding studies in normal healthy volunteers or patient studies (e.g. for oncology compounds) should (Patterson et al., 2000d; see Lesko et al., 2000 for further discussion):

> "1. Provide the range of safe (and potentially efficacious) doses for further study in patients,
> 2. Provide an initial description of pharmacokinetic exposure levels and/or biomarker/surrogate marker levels at each dose to facilitate choice of dose, dose titration, and dosing interval in Phase II studies,
> 3. Develop initial models for use in pharmacokinetic-pharmacodynamic modeling for both desirable and undesirable effects."

Subsequent Phase II clinical studies in patients establish the minimum starting and maximum effective dose as well as the maximum tolerated dose in patients with the disease state using pharmacodynamic endpoints or surrogate markers of therapeutic response. Dose titration and the length of time needed to see an effect (desirable or undesirable) are also established. In these studies, models relating dose to PK and to PD are developed to understand the mechanism of the drug's action and to search for relevant covariates (e.g. age or gender) to control later Phase II or Phase III confirmatory trial designs (International Conference on Harmonization Guidance ICH-E4, 1994). Dose finding studies in Phase II studies in the target population

should (Patterson et al., 2000d):

> "1. Establish the therapeutic window by identifying a minimum effective starting dose (the lowest dose yielding a desirable effect), a maximum effective dose (the dose beyond which further escalation lacks further desirable benefit), and a maximum tolerated dose (the dose beyond which there is an unacceptable increase in undesirable effects) in the target population,
> 2. Identify the time interval needed to see an effect (desirable and/or undesirable) and reasonable, response-guided titration steps along with the time intervals at which to dose titrate,
> 3. Develop updated pharmacokinetic-pharmacodynamic models for both desirable and undesirable effects in the population of interest, and identify potential covariates to be studied for dose adjustment in Phase III (e.g. age, gender)."

Once a dose or set of efficacious doses are chosen from Phase II trials, confirmatory trials are subsequently performed to support regulatory acceptance. These trials, in large numbers of patients with the disease under study, should characterize the risk relative to benefit in clinical use of the compound. These studies in late Phase II or Phase III should be used to (Patterson et al., 2000d):

> "1. Establish the risk:benefit ratio and pharmacokinetic-pharmacodynamic relationship (if any) for doses chosen to be in the therapeutic window established in Phase II,
> 2. Finalize pharmacokinetic-pharmacodynamic models developed in Phase II, and identify adjustments to dosing procedures appropriate to special populations for the drug under study."

Generally, from the time a drug enters the clinic to the time it is approved and ready to market 10.4 years on average elapse (DiMasi, 2001). Many approaches to speeding this process have been considered (see for example, Patterson, 2002i). We will not develop these approaches further but instead will dwell on bridging approaches using pharmacokinetics for the purpose of speeding global market access as part of clinical development planning using an ICH-E5 based approach.

We will concentrate on obtaining a firm understanding of the comparison between populations in pharmacokinetics early in the development process. Assuming a first-time-in-human study has been performed in Western volunteers and that a maximum tolerated dose has been identified, it should, theoretically, be possible with the additional support of in vitro and preclinical data to evaluate the nine conditions described above to determine a compound's comparative ethnic sensitivity.

Historically, however, pre-clinical and in vitro evidence has sometimes been proven unreliable in drug development (DiMasi, 2001), and it is desirable to explore findings from this exercise using a small, randomised, single-dose, placebo-controlled parallel group PK study in an ethnic group different from those studied in the USA or Europe. In early phase planning, to characterize potential differences in PK across doses and populations, another option is to perform a single dose, dose ranging, placebo controlled, parallel group PK bridging study in Pan-Asian volunteers as part of late Phase I or early Phase II studies.

Often Japanese volunteers are studied for this purpose and PK exposure levels are compared back to Western volunteers as there is a significant population of such volunteers in Australia and Hawaii. Alternatively, a South Korean or Taiwanese population in Pan-Asia may be utilised as these nations are to some degree westernised. Exposure (AUC, Cmax) pharmacokinetic data are summarised in both populations. The objective of the study is to estimate the difference (if any) in extent (AUC) and rate (Cmax) of exposure between the populations.

More than one such bridging study may be required for different markets or regulatory requirements depending upon the properties of the drug product - it is not yet known whether differences in intrinsic/extrinsic factors within the Pan-Asian market will require more than one such PK study (Aarons et al., 2001). ICH E5 discusses three ethnic groups: Asians, Caucasians and Blacks. Some previous reports (Naito, 1994a-b; Yusahura, 1994) had claimed that 'inter-ethnic differences were no larger than intra-ethnic variation' in pharmacokinetics for most medicines implying that statistically significant changes could be unusual. This view is not widely held and is at odds with the extensive debate on population and individual bioequivalence (see Chapter 1) and is not in accord with the practices of most regulatory agencies with respect to the considerations in population pharmacokinetics (for example FDA Guidance, 1999c). This finding also does not agree with recent reports on the impact of inter-ethnic intestinal CYP3A metabolism and genetic polymorphism (Mancinelli et al., 2001) and would appear to be misleading.

These Pan-Asian PK studies should characterize the PK-exposure levels across a range of doses comparable to those studied in the Western Phase I studies or alternatively, at the maximum tolerated dose. Placebo should be used in order to ensure that any unexpected safety findings are placed in context.

These bridging studies should be regarded as exploratory in nature and as part of the 'learning' process enhancing the information-base for subsequent, more rigorous development (Sheiner, 1997; Sheiner and Steimer, 2000). In combination, the full information from the complete Phase I Western clinical development program and the Pan-Asian PK-bridging study should provide sufficient information to evaluate whether any gross differences between populations in pharmacokinetics could result in qualitative changes to PD response.

This constitutes a shift in focus relative to the techniques utilised in Chapters 1 through 5 of this thesis in that statistical analysis will be exploratory rather than confirmatory, though many of the concepts developed earlier will be helpful. Statistical methods for study design and approaches to analysis of pharmacokinetic data arising from such studies will be developed in this thesis. The approach taken to analysis will be exploratory rather than confirmatory, and we will refer to such studies as BBE ('Bridging, Bioequivalence style') studies.

For some few products which are insensitive to intrinsic and extrinsic factors, and where PK can serve as a suitable surrogate or intermediate marker (Sheiner, 1997) for safety and efficacy, a PK only bridging strategy may be sufficient to secure market access (though we expect such products to be rare, given the discussion of this Chapter). We will develop sample size requirements using simulation to assess how to use the metrics to establish a claim, if such an extension to the exploratory methods developed in this thesis is desirable. We will subsequently discuss whether such an approach is reasonable.

Some methods have been described previously (Liu, 2000; Kawai et al., 2000; Sarkar et al., 2002), but these will not be developed for application in this thesis as they relate primarily to clinical or pharmacodynamic data.

Although much progress has been made since ICH-E5 became effective in 1998 there is still much diversity regarding how the guidelines are interpreted by the various countries within the Pan-Asian area. Currently Japan (Naito, 1998b) is considered to take a far more conservative approach to bridging than other countries such as Taiwan (EFPIA-99, 1999) or South Korea. Under certain conditions, pharmacokinetic bridging is acceptable in South Korea and Taiwan. In general however, at least one dose response study (as detailed in ICH-E4, 1994), and in many situations a Phase III trial, in Japanese patients is required for regulatory approval in Japan (though recent reports such as Nagata et al., 2000 indicate that this policy may be weakening).

It should be realised that in regions where there is little experience with registration based on clinical data from the original region, the regulatory authorities may, prior to approval, request a bridging study even for compounds insensitive to ethnic factors.

We will now discuss concepts in study design important for consideration in BBE studies.

6.3 Topics in Bridging Bioequivalence PK Study Design

We will consider BBE designs with a test and reference population where one dose level (a single administration) is of interest (for a single formulation of drug product).

In a BBE study design, the classic factors involved in experimental design (randomisation and blinding, replication, and blocking; Hinkelmann and Kempthorne, 1994) should be considered.

Randomisation and Blinding: It is recommended that normal healthy volunteer subjects (the experimental unit) within the test (the population in the new region) and reference (the population in the original region) populations be randomly assigned to receive a dose of drug product or placebo, so that, if unexpected adverse events are observed in either population, these can be placed in context of spontaneously occurring events. In consideration of such factors, it is recommended that both subjects and clinical personnel be blinded to placebo or drug administration (i.e. double blinded). In general, as these studies will be performed early in drug development, it may be the case that enrollment will be restricted to male subjects only, until definitive repro-toxicology results are available. If female volunteers can be included in the study, randomisation should be stratified by gender within each population.

Choice of dose in the test population should be carefully considered and based upon what was observed in the reference population. Allometric scaling techniques using population pharmacokinetic modelling (see Reigner and Blesch, 2001) may also be considered when choosing the dose (or doses) to be used in the new population.

Replication: BBE studies where pharmacokinetics are the endpoints of interest are likely to be conducted early in drug development (see Section 6.2), and in this context, it is likely that only few subjects ($n = 8 - 12$) in each population will be exposed to drug in order to ensure that, if the drug is unsafe, few subjects will be exposed to risk. This is in line with recent communications summarising the ethics of dosing normal healthy volunteers with experimental drug substance, regardless of national or regional considerations (Tanida, 2002). It is assumed

that subjects will not derive any medical benefit from the study. Study design then consists of careful consideration of the intrinsic an extrinsic factors to ensure an unbiased comparison of populations is captured.

In general, at the point in drug development at which we consider such trials will be employed, it will generally be the case that reliable estimates of between-subject and of within-subject variation in the reference population will be available. We propose that sample size be selected in the new population (if safety issues are not of great concern) in order to provide a given level of precision in the study findings relative to the original population. We note here that in general, drug development will be in such an early stage that no pre-specified difference in population means will be of interest nor will testing whether a difference relative to a goalpost of nature similar to those utilised in bioequivalence be of interest.

The 'estimation' approach can be useful when the magnitude of effect is not known and the main study objective is to provide evidence of what the potential value, or range of values, may be, or when the sample size is in part set by feasability, and we wish to provide an idea of the precision the trial is likely to provide for the effect of interest. In such cases, the intent is to provide an estimate of the expected width or precision of the plausible range of values as expressed by a confidence interval. This will help satisfy our expectation with regard to acceptability and applicability of study results in the knowledge that, 'The confidence interval can be thought of as the set of true but unknown differences that are statistically compatible with the observed difference.' (Goodman, 1994)

First consider AUC and Cmax data from a parallel group design where independent unbiased method-of-moments estimators $\hat{\delta}$, $\hat{\sigma}_T^2$, and $\hat{\sigma}_R^2$ are derived for $\delta = \mu_T - \mu_R$, σ_T^2, and σ_R^2. AUC and Cmax pharmacokinetic data are log_e-normally distributed (see Chapter 1), and are sufficiently described by mean μ_t and variance σ_t^2, where t represents test and reference populations.

Where intrinsic and extrinsic factors are sufficiently precluded to ensure that comparison between populations is not confounded, traditional method-of-moment techniques (Rao, 1973) may be used (following log_e-transformation) to derive unbiased estimators $\hat{\mu}_t$ and $\hat{\sigma}_t^2$ for the moments of interest. Let Y_{ij} be the log_e-transformed AUC or Cmax for subject j $(j = 1, \ldots, n_i)$

of population i (i =Test, Reference)

$$\hat{\mu}_t = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i}$$

and

$$\hat{\sigma}_t^2 = \frac{\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{n_i - 1}$$

Here then

$$\hat{\delta}^2 \sim \left(\frac{\sigma_T^2}{n_T} + \frac{\sigma_R^2}{n_R}\right) \chi_1^{2'} \left(\frac{\delta^2}{\frac{\sigma_T^2}{n_T} + \frac{\sigma_R^2}{n_R}}\right)$$

where $\chi^{2'}\left(\frac{\delta^2}{\frac{\sigma_T^2}{n_T}+\frac{\sigma_R^2}{n_R}}\right)$ represents the non-central chi-squared distribution with non-centrality parameter $(\frac{\delta^2}{\frac{\sigma_T^2}{n_T}+\frac{\sigma_R^2}{n_R}})$,

where

$$\hat{\sigma}_T^2 \sim \frac{\sigma_T^2(\chi_{\nu_T}^2)}{\nu_T}$$

with $\chi_{\nu_T}^2$ representing the central chi-squared distribution with $\nu_T = n_T - 1$ degrees of freedom,

and

where

$$\hat{\sigma}_R^2 \sim \frac{\sigma_R^2(\chi_{\nu_R}^2)}{\nu_R}$$

with $\chi_{\nu_R}^2$ representing the central chi-squared distribution with $\nu_R = n_R - 1$ degrees of freedom.

In the context of BBE in this Chapter, the subscripts T and R refer to test and reference populations, respectively. In such a design, $\hat{\delta}$, $\hat{\sigma}_T^2$, and $\hat{\sigma}_R^2$ are independent and unbiased estimators.

Then $\hat{\delta} = (\hat{\mu}_T - \hat{\mu}_R) \sim N(\mu_T - \mu_R, (\sigma_T^2 + \sigma_R^2)/n))$ where n is the sample size (per group) in a two-group, balanced, parallel design with T=Test and R=Reference *populations* in this context where N represents the normal distribution with (mean, variance). Further $\hat{\delta} = (\hat{\mu}_T - \hat{\mu}_R) \sim N(\mu_T - \mu_R, (\frac{\sigma_T^2}{n_T} + \frac{\sigma_R^2}{n_R}))$ where n_i is the sample size (in population $i = T, R$ for T =Test and R =Reference populations) in a two-group parallel design where sample size is not presumed equal. Then, a 90% confidence interval for $\mu_T - \mu_R$ is:

$$\hat{\delta} \mp t_{\hat{\nu}}(0.95) \sqrt{\frac{\hat{\sigma}_T^2}{n_T} + \frac{\hat{\sigma}_R^2}{n_R}}$$

where $t_{\hat{\nu}}(0.95)$ is the 95th quantile of a t-distribution with Satterthwaite (1941) degrees of freedom

$$\hat{\nu} = \frac{(\sum_i(\hat{\sigma}_i^2)/n_i)^2}{\sum_i(\hat{\sigma}_i^4/n_i^2(n_i-1))}$$

Sample size may then be derived using the techniques described previously in this thesis to determine the necessary sample size to conclude equivalent exposure. Alternatively, the nonparametric percentile bootstrap procedure (Efron and Tibshirani, 1993) may be used to construct a confidence interval, and we will assess which is most appropriate later in this Chapter.

Consider

$$w_\delta = t_{\hat{\nu}}(0.95)\sqrt{\frac{\hat{\sigma}_T^2}{n_T} + \frac{\hat{\sigma}_R^2}{n_R}}$$

This function provides a 'precision' estimate for the true mean difference. Goodman (1994) notes that use of a method like that proposed above should be exercised with caution as, in a situation where the study design is truly intended to support a test of hypothesis, the approach corresponds to a test using 50% power when precision is equal to the difference of interest. Similarly, in situations where an equivalence approach is intended (i.e. 90% CI are calculated), the method presented in this thesis corresponds to a two-one sided hypothesis test with 50% power when precision is equal to the equivalence range of interest.

When a pre-specified goalpost is available for utilisation in equivalence testing, it is easy to construct a two-one sided testing procedure akin to that used in average bioequivalence testing. Here, let $x > 0$ be a pre-defined scalar such that differences in populations in excess of x are undesirable. Then we will test the following two one-sided tests to evaluate equivalence in population means:

$$H_{01} : \mu_T - \mu_R \geq x$$

$$H_{02} : \mu_T - \mu_R \leq -x$$

However this type of 'bioequivalence' approach to choice of sample size is dependent upon the variation and the true difference in populations (see Senn, 1997); both of which are unlikely to have been studied at this stage of drug development. Later in this Chapter, we will also consider some additional metrics for BBE assessment. These metrics will incorporate the population variances in a manner similar to that considered in PBE testing.

Consider now a comparison of the variances between populations. As above we know that $\hat{\sigma}_T^2 \sim \sigma_T^2 \frac{\chi^2_{n_T-1}}{n_T-1}$ where $n_T - 1$ is the degrees of freedom associated with $\hat{\sigma}_T^2$, and $\hat{\sigma}_R^2 \sim \sigma_R^2 \frac{\chi^2_{n_R-1}}{n_R-1}$ where $n_R - 1$ is the degrees of freedom associated with $\hat{\sigma}_R^2$ where these two estimates of variance are independent. Here a 90% confidence interval for $\frac{\sigma_T^2}{\sigma_R^2}$ will be:

$$(\frac{\hat{\sigma}_T^2}{\hat{\sigma}_R^2} F_{n_R-1,n_T-1}(0.05), \frac{\hat{\sigma}_T^2}{\hat{\sigma}_R^2} F_{n_R-1,n_T-1}(0.95))$$

such that $w_{\sigma_T^2/\sigma_R^2} = max(F_{n_R-1,n_T-1}(0.05), F_{n_R-1,n_T-1}(0.95))$

$= F_{n_R-1,n_T-1}(0.95)$ where $F_{n_R-1,n_T-1}(\alpha)$ denotes the αth percentage of the F-distribution (Muirhead, 1982).

Here also, when a pre-specified goalpost is available for utilisation in equivalence testing, it is easy to construct a two-one sided testing procedure akin to that used in average bioequivalence testing. Here, let $y > 0$ be a pre-defined scalar such that changes in variance in populations in excess of y are undesirable. Then we will test the following two one-sided tests to evaluate equivalence in population variances:

$$H_{01} : \frac{\sigma_T^2}{\sigma_R^2} \geq y$$

$$H_{02} : \frac{\sigma_T^2}{\sigma_R^2} \leq 1/y$$

In the example to be considered later in this Chapter, $\sigma_R^2 = 0.0632$ for $n_R = 53$ caucasian male subjects dosed with an innovator drug product. The intent of the study to be planned was to estimate the difference in AUC and Cmax in a new South Korean, male, healthy volunteer population, who previously had not been exposed to drug. Only a few Korean subjects would be initially exposed, in order to ensure that few subjects were exposed to potential risk. The drug product in question was expected to be ethnically insensitive based upon the pharmacokinetic profile in caucasians.

An initial sample size of $n_T = 8$ was proposed. Based on the above estimate of variation for AUC and assuming homogeneity in variation between populations, we find that $\hat{\nu} = 9.24$ and $w_\delta = 0.1742$. Thus we would expect the half-width of a confidence interval for the difference in mean AUC to be approximately 19% of the point estimate; certainly acceptable for the purposes of studying the pharmacokinetics between populations.

However, the precision associated with the ratio of variances is $F_{52,7}(0.95) = 3.31$. This is informative in that this finding indicates that it is unlikely we will be able to make any meaningful inference between population variances. To increase precision, n_T must be increased; however, it was judged sufficient that precision be adequate for comparison of mean AUC and that only extreme differences in variation be recognisable.

Precision in other metrics we will consider later in this chapter are less straightforward (especially when the poor precision of the comparison of variances is taken into account). While the mean and variance of these expressions are readily derivable, in studies with such small sample sizes, they are not likely to aid in defining a precision estimate. Nor would such an estimate be readily interpretable in the context of a composite metric based on the combination of means and variances. We will instead develop a nonparametric bootstrap program and combine it in application with a simulation mechanism to address precision in study planning for such metrics. The power of such metrics to demonstrate BBE will be studied using simulation and compared to several potential goalposts.

Blocking: To isolate differences between populations and ensure unbiased comparison of exposure levels, intrinsic and extrinsic factors should be controlled. Gender stratification (see Miller, 2001) and age, height, and weight ranges should be made as homogeneous as possible between populations. Subjects should be in good physical condition and should not have any concomitant disease state which would impact exposure levels and confound interpretation (e.g. liver, kidney, or cardiovascular disease). Subjects should not be taking any concomitant medication (including smoking and alcohol) and should be on a consistent diet and exposed to the same set of climatic variables (e.g. light, exercise, stress levels) while being studied.

Pharmacokinetic properties of the drug product should be carefully considered when selecting subjects for participation in the study. For example, in drugs known to be metabolised by certain pathways in the liver (CYP2D6), in Caucasians, there is known to be a rare sub-population (some 10% of the population) in which subjects metabolise a drug metabolised by this pathway slowly. Thus in this 'poor-metaboliser' population, exposure levels of the parent compound will be dramatically increased relative to the rest of the population. If a drug is known to be prone to such ADME properties (see Chapter 1), subjects should be screened and randomised to ensure that the number of volunteers with such a genetic polymorphism is homogeneous between

populations.

Standardisation of the application of study procedures to the experimental units should be ensured. Clinical practice should be standardised and conducted in accordance with good clinical practice (e.g. time and content of meals, time and administration of and compliance with study medication), and the pharmacokinetic (and other) sampling schemes should be standardised across populations. Clinical personnel should ensure that subjects receive study medication in a standard manner (e.g. after an overnight fast) according to the randomisation schedule.

Of particular importance in BBE studies involving pharmacokinetics is the quality and uniformity of the assay and shipping of the biological materials to the site of assay. Assays used should be validated and if performed at different sites should be confirmed to provide equivalent results (to do otherwise risks biasing the study results or at the least increasing noise). Optimally, to avoid these issues, shipping procedures should be arranged to ensure that, regardless of point of origin, samples do not degrade and arrive at the single site of assay in sufficient time to be utilised under good laboratory practices.

Under some circumstances, it will not be possible to ensure that all intrinsic and extrinsic factors are homogeneous across populations (e.g. weight in Pan-Asian volunteers is unlikely to be homogeneous with Western volunteers). Under such circumstances, the factor involved should be carefully controlled in a manner appropriate to the eventual patient population to which the drug will be applied, and the model to be used in analysis should account for this factor.

One such factor, deserving of special consideration, is dose. If knowledge of the dose to PK response relationship is not well established by the Phase I studies in the original region, a dose ranging design in the BBE study may be employed. Under such a design, subjects would be randomly assigned to placebo or one of several active doses of drug. Dose then would be one of the factors explored in detail in such a model.

We now turn to the consideration of data from a single dose comparison.

6.4 Single Dose, Placebo Controlled BBE Studies

6.4.1 Background

For the purposes of this thesis, it is assumed that a single active dose, randomised (to placebo control), clinical trial is conducted for the purposes of estimating the difference in primary endpoints AUC and Cmax (from the subjects receiving active regimen) between the test and reference populations. For those few compounds known to be ethnically insensitive (see previous discussion in this chapter), it may be the case that the study is performed to establish that the populations are equivalent in terms of their AUC and Cmax, and we now turn to the consideration of metrics to be used in such measurement.

6.4.2 Metrics

It was originally proposed that population bioequivalence be assessed using the following aggregate statistic (FDA Guidance, 1997).

$$\frac{(\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2}{max(0.04, \sigma_R^2)} \tag{65}$$

where $\sigma_T^2 = \sigma_{WT}^2 + \sigma_{BT}^2$ and $\sigma_R^2 = \sigma_{WR}^2 + \sigma_{BR}^2$. Note that this aggregate statistic can be constructed using a mixed model from a parallel group, two-period cross-over design, or other (e.g. replicate) design. We refer to this metric as *constant*-scaled if $\sigma_R^2 < 0.04$; otherwise the metric is deemed to be *reference*-scaled as the variance for the reference formulation appears in the denominator. Constant-scaling was introduced (see Chapter 1) as a means of keeping low-variability products from being held to what was felt to be an unreasonable strict standard for bioequivalence.

For the purposes of Bridging bioequivalence, we will neglect constant-scaling and consider only the reference scaled metric.

$$\frac{(\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2}{\sigma_R^2} \tag{66}$$

Note that, due to the nature of this 'aggregate' criterion, differences in means in this criterion can be 'negated' by decreased variance for the test formulation. As discussed in previous Chapters, some have noted this to be an undesirable property of the proposed metric, (Endrenyi and Hao, 1998c), and it is known that such trade-offs do occur in practice in cross-over trials (Zariffa et al., 2000).

The goalpost for population bioequivalence (see Chapter 4) assessment assumes the variance for the reference formulation is 0.04. The difference $\mu_T - \mu_R$ is allowed to take on a value of up to 0.2231 and a variance allowance of 0.5 in the numerator under the procedure proposed by the FDA (cf. FDA Guidance, 1999) and (66). Thus the regulatory 'cut-off' is constrained to a level of 1.74. If the upper 95% bound on the FDA metric falls below this value of 1.74, population bioequivalence is demonstrated for the endpoint under study. Scaled to reference variation (again, assumed to be 0.04, under the FDA Guidance 1997), the goalpost accounting for the means amounts to a value of $1.24=(log_e(1.25))^2/0.04$. The remaining allowance, known as the 'variance allowance' and is equal to 0.5 (allowing for a difference in variances $((\sigma_T^2 - \sigma_R^2)/\sigma_R^2)=(0.02/0.04)$, when scaled to reference product variation.)

This is not an appropriate choice of goalpost when looking at a BBE study. Obviously, this is an exploratory study, so one would debate whether a 'goalpost' is at all relevant. However, one might set a number beyond which it would be extremely undesirable to see any potential for responses in the limited data such an exploratory PK study would generate. We will illustrate how to do so separating the metric into components $x = \frac{(\mu_T - \mu_R)^2}{\sigma_R^2}$ and $y = \frac{\sigma_T^2}{\sigma_R^2}$. Note that (66) is equal to $x + y - 1$.

In Figure 42, the response-surface for combinations of x and y yielding (66) is plotted both as a surface and, in the second part of the figure, as a projection onto the plane of possible responses as a function of x and y. Here we see the potential for trade-offs in (66). For example, for an (66) value equal to 1, it is observed that x can be as large as 2 if the variation in the test population is very small relative to reference. Equivalently, we can see such a value when the ratio of test to reference variances is 2, but x is near 0.

As discussed previously in Chapter 4, this potentially helps explain the findings of Chapter 4 and 5 in that PBE is so liberal as to allow nearly any product market access for a large sample size.

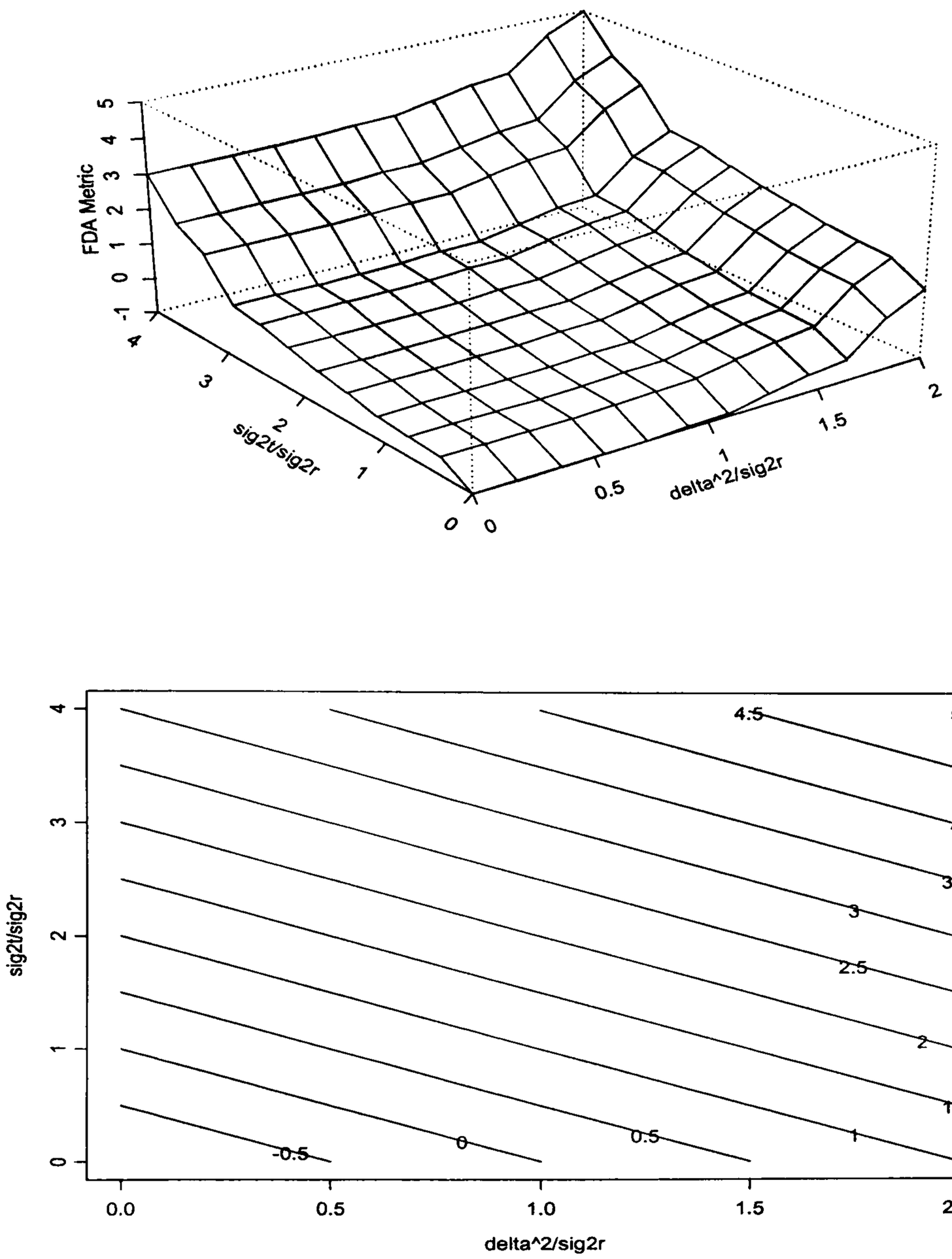


Figure 42: Response-Surface and Projected Values of (66)$= x + y - 1$ relative to $x = \frac{(\mu_T - \mu_R)^2}{\sigma_R^2}$ and $y = \frac{\sigma_T^2}{\sigma_R^2}$

Dragalin and Fedorov (1999a) introduced an alternative discrepancy measure for measuring the divergence in two independent distributions based on measures of Kullback-Leibler distance (Kullback, 1968). In the situation where two parallel groups are being assessed, this measure is as follows:

$$d(f_T, f_R) = \frac{1}{2}\left\{(\mu_T - \mu_R)^2 + \sigma_T^2 + \sigma_R^2\right\}\left(\frac{1}{\sigma_T^2} + \frac{1}{\sigma_R^2}\right) - 2 \tag{67}$$

Here, we see a metric which does not allow 'rewards' for decreases in variance to offset changes in means. In fact (67) is equal to 0 if and only if $\mu_T - \mu_R = 0$ and $\sigma_T^2 = \sigma_R^2$. Separating the metric into components $x = \frac{(\mu_T - \mu_R)^2}{2\sigma_R^2} + \frac{(\mu_T - \mu_R)^2}{2\sigma_T^2}$ and $y = \frac{\sigma_T^2}{2\sigma_R^2} + \frac{\sigma_R^2}{2\sigma_T^2}$. Note that (67) is equal to $x + y - 1$. Here again, we see trade-offs in terms of varying the x and y space to achieve a given value of (67). See Figure 43.

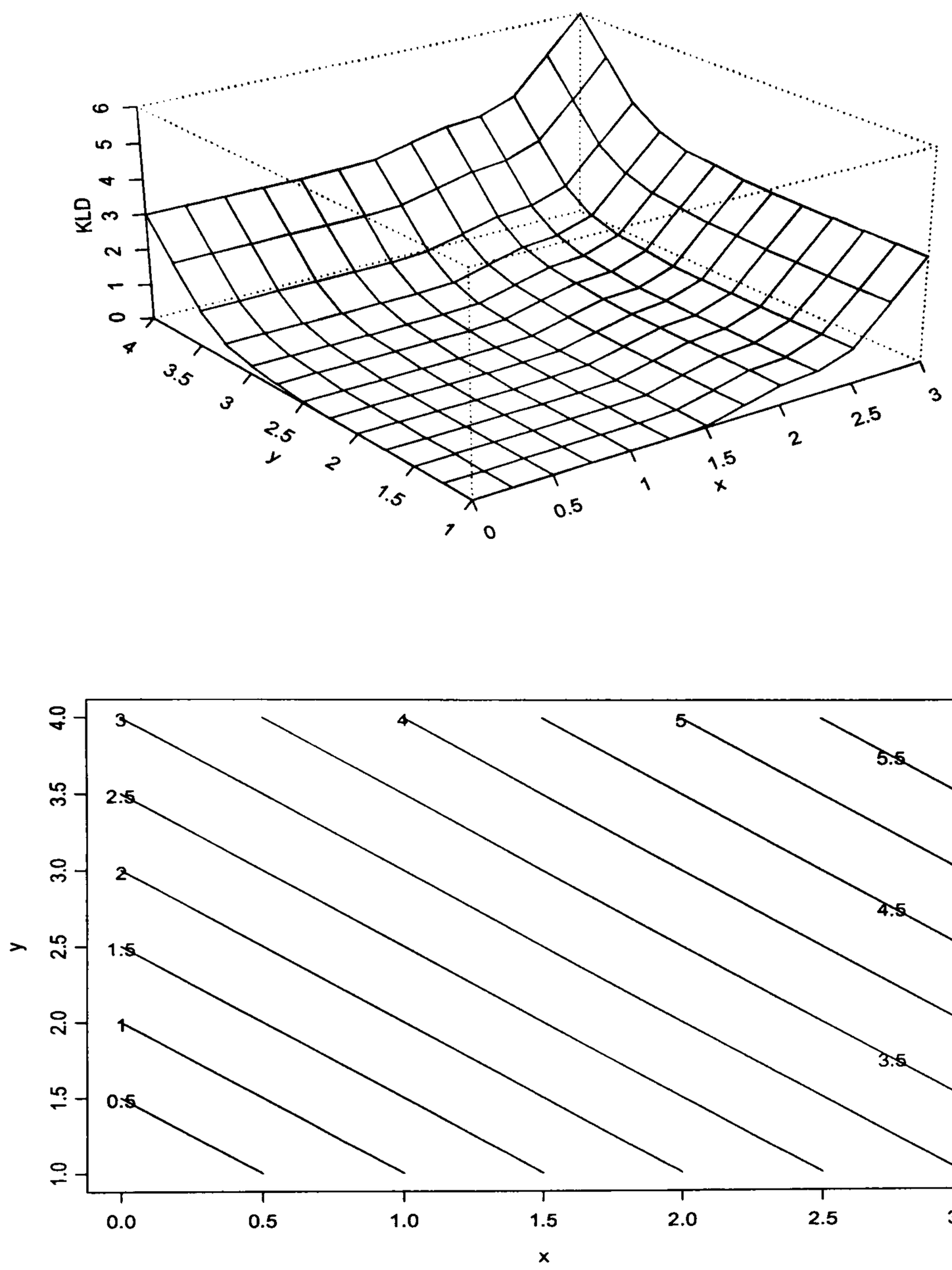


Figure 43: Response-Surface and Projected Values of (67)$= x + y - 1$ relative to $x = \frac{(\mu_T - \mu_R)^2}{2\sigma_R^2} + \frac{(\mu_T - \mu_R)^2}{2\sigma_T^2}$ and $y = \frac{\sigma_T^2}{2\sigma_R^2} + \frac{\sigma_R^2}{2\sigma_T^2}$

Finally, one might compare the estimates for the means and variances themselves using established methods for their comparison and evaluate them relative to a goalpost to assess clinical relevance. We will explore this approach in the subsequent sections also.

6.4.3 Estimation and Properties of Estimates

The estimation procedures follow the general principles of a 'plug-in' method. Estimates for the parameters of interest (δ, σ_T^2, σ_R^2) are derived using method-of-moments or restricted maximum likelihood estimation. These estimates ($\hat{\delta}$, $\hat{\sigma}_T^2$, $\hat{\sigma}_R^2$) and then analysed in accordance with the procedures described above (for $\hat{\delta}$ and $\frac{\hat{\sigma}_T^2}{\hat{\sigma}_R^2}$ or 'plugged-in' to the formulas for the FDA and KLD metrics to provide an estimate for the metrics.

In this context, much like PBE, the statistics $\hat{\mu}_t$ and $\hat{\sigma}_t^2$ accounting for n_t may be compared in some fashion to determine whether population exposure levels are sufficiently similar to be termed 'bioequivalent' (see Chapter 1 and Chapter 2.). However, unlike PBE, in BBE studies, intrinsic and extrinsic factors may confound inference. A REML model based approach, accounting for intrinsic and extrinsic factors potentially can be used to derive unbiased estimates for the moments of interest where factors are known to differ. Such an example will be considered. In such a study, as is common practice, it is assumed that subjects are independent. We will consider $SAS^{©}$ based REML procedures which allow for the inclusion of continuous independent variables while allowing for the differentiation of variances between populations.

The REML procedure to be considered does not constrain the variance to be homogenous across populations. As known from Searle (1971) estimates resulting from this model should equate to a method-of-moments approach in complete and strongly balanced data sets based on $SAS^{©}$ code as follows:

```
PROC MIXED METHOD=REML SCORING=50 MAXITER=200;
CLASS POP;
MODEL lnAUC = POP covariates/DDFM=KENWARDROGER;
REPEATED /group=POP;
LSMEANS POP;
```

Note that beginning here, we will adopt the use of the Kenward-Roger (1997) degrees of freedom as implemented in $SAS^{©}$ as this procedure also inflated the variance estimate for δ to account for the iterative nature of REML estimation according to the approach of Harville and

Jeske (1992). Satterthwaite's (1941) procedure, as currently implemented in $SAS©$, does not account for this and should be expected to provide a slightly lower level of coverage probability.

The metric proposed by the FDA (FDA Guidance, 1997, 1999a, 1999b, 2000b) for the assessment of population bioequivalence under such a design is asymptotically unbiased, though in small samples it carries a small positive bias also related to degrees of freedom. We will also show how an unbiased estimator can be derived.

We now consider the bias in the FDA metric for such a design.

Theorem 6.1 *Bias in the FDA Metric for BBE*

A method of moments estimator for the FDA metric for assessment of BBE is

$$\frac{\hat{\delta}^2 + \hat{\sigma}_T^2 - \hat{\sigma}_R^2}{\hat{\sigma}_R^2} = \frac{\hat{\delta}^2}{\hat{\sigma}_R^2} + \frac{\hat{\sigma}_T^2}{\hat{\sigma}_R^2} - 1 \tag{68}$$

The expected value of expressions (68) is asymptotically unbiased but is positively biased in small samples.

Proof: Taking expectations and assuming independence in (68),

$$E\Big(\frac{\hat{\delta}^2}{\hat{\sigma}_R^2} + \frac{\hat{\sigma}_T^2}{\hat{\sigma}_R^2} - 1\Big)$$

$$= E\frac{\frac{\sigma_T^2+\sigma_R^2}{n}\big(\frac{\hat{\delta}^2}{(\sigma_T^2+\sigma_R^2)/n}\big)}{\sigma_R^2\big(\frac{\hat{\sigma}_R^2}{\sigma_R^2}\big)} + E\frac{\sigma_T^2\big(\frac{\hat{\sigma}_T^2}{\sigma_T^2}\big)}{\sigma_R^2\big(\frac{\hat{\sigma}_R^2}{\sigma_R^2}\big)} - 1$$

Further, using the results of Muirhead (p 24-25, 1982), it is seen that this expression reduces to,

$$\frac{n-1}{n-3}\Big[\frac{\delta^2}{\sigma_R^2} + \frac{(n+1)\sigma_T^2}{n\sigma_R^2} + \frac{1}{n}\Big] - 1 \tag{69}$$

As sample size increases this expression becomes,

$$lim_{n\to\infty}\frac{n-1}{n-3}\Big[\frac{\delta^2}{\sigma_R^2} + \frac{(n+1)\sigma_T^2}{n\sigma_R^2} + \frac{1}{n}\Big] - 1 = \frac{\delta^2}{\sigma_R^2} + \frac{\sigma_T^2}{\sigma_R^2} - 1$$

which is an unbiased estimate for (66). However in small samples, the bias is

$$\Big(\frac{n-1}{n-3}-1\Big)\frac{\delta^2}{\sigma_R^2}+\Big(\frac{(n-1)(n+1)}{n(n-3)}-1\Big)\frac{\sigma_T^2}{\sigma_R^2}+\frac{1}{n}\geq 0$$

□□□ Thus the estimation procedure is positively biased (against sponsors) when using a 'plug-in', method-of-moments estimation procedure. An unbiased estimator may be derived as follows:

Theorem 6.2 *An Unbiased FDA Metric for BBE*

An unbiased method of moments estimator for the FDA Metric in BBE is

$$\frac{n-3}{n-1}\Big(\frac{\hat{\delta}^2}{\hat{\sigma}_R^2}+\frac{(n-1)\hat{\sigma}_T^2}{n\hat{\sigma}_R^2}\Big)-1-\frac{1}{n} \tag{70}$$

Proof: Taking expectations and assuming independence in (70),

$$=\frac{n-3}{n-1}\Big(E\frac{\frac{\sigma_T^2+\sigma_R^2}{n}\big(\frac{\hat{\delta}^2}{(\sigma_T^2+\sigma_R^2)/n}\big)}{\sigma_R^2\big(\frac{\hat{\sigma}_R^2}{\sigma_R^2}\big)}+E\frac{(n-1)\sigma_T^2\big(\frac{\hat{\sigma}_T^2}{\sigma_T^2}\big)}{n\sigma_R^2\big(\frac{\hat{\sigma}_R^2}{\sigma_R^2}\big)}\Big)-1-\frac{1}{n}$$

Further, using the results of Muirhead (p 24-25, 1982), it is seen that this expression reduces to,

$$\Big[\frac{\delta^2}{\sigma_R^2}+\frac{\sigma_T^2}{\sigma_R^2}\Big]-1$$

□□□

We now develop an extension to the above proofs when sample size differs between populations.

Theorem 6.3 *Bias in the FDA Metric for BBE when Sample Size Differs between Populations*

When a method-of-moments estimator for the FDA metric is derived as in (68) where n_T and n_R are the sample size in test and reference populations, then the expected value is

$$\frac{\nu_R}{\nu_R-2}\Big[\frac{\delta^2}{\sigma_R^2}+\frac{(n_T+1)\sigma_T^2}{(n_T)\sigma_R^2}+\frac{1}{n_R}\Big]-1 \tag{71}$$

This expression (71) is asymptotically unbiased for (66), but is positively biased by

$$\frac{2\delta^2}{\sigma_R^2(n_R-3)}+(\frac{2+n_R/n_T-1/n_T}{n_R-3})\frac{\sigma_T^2}{\sigma_R^2}+\frac{1-1/n_R}{n_R-3}$$

An unbiased estimator is

$$\frac{\nu_R-2}{\nu_R}\Big[\frac{\hat{\delta}^2}{\hat{\sigma}_R^2}+\frac{(n_T-1)\hat{\sigma}_T^2}{(n_T)\hat{\sigma}_R^2}-\frac{\nu_R}{(\nu_R-2)n_R}\Big]-1 \tag{72}$$

where ν_t is the degrees of freedom associated with $\hat{\sigma}_t^2$ for t =test and reference populations.

Proof: Taking expectations of (68) and using the results of Muirhead (p 24-25, 1982), it is seen that this expression reduces to:

$$\Big(\frac{\frac{\sigma_R^2}{n_R}+\frac{\sigma_T^2}{n_T}}{\sigma_R^2}\Big)F'+\frac{\sigma_T^2}{\sigma_R^2}F-1$$

where F' is a random variable such that $F' \sim F'_{1,\nu_R}\Big(\frac{\delta^2}{\frac{\sigma_R^2}{n_R}+\frac{\sigma_T^2}{n_T}}\Big)$ and F is a random variable such that $F \sim F_{\nu_T,\nu_R}$. Taking the expectation, the expression (71) is found. As sample size increases (71) becomes,

$$lim_{n_R\to\infty,n_T\to\infty}\frac{\nu_R}{\nu_R-2}\Big[\frac{\delta^2}{\sigma_R^2}+\frac{(n_T+1)\sigma_T^2}{(n_T)\sigma_R^2}+\frac{1}{n_R}\Big]-1=\frac{\delta^2}{\sigma_R^2}+\frac{\sigma_T^2}{\sigma_R^2}-1$$

which is asymptotically unbiased for (66). Consider (72), which can be expressed as

$$\frac{\nu_R-2}{\nu_R}\Big[\Big(\frac{\frac{\sigma_R^2}{n_R}+\frac{\sigma_T^2}{n_T}}{\sigma_R^2}\Big)F'+\frac{(n_T-1)\sigma_T^2}{(n_T)\sigma_R^2}F-\frac{\nu_R}{(\nu_R-2)n_R}\Big]-1$$

Taking the expectation, this expression reduces to

$$\frac{\delta^2}{\sigma_R^2}+\frac{\sigma_T^2}{\sigma_R^2}-1$$

□□□

Note the metric is positively biased (against sponsors) in this setting though the expected bias is small and negligible due to the sample sizes expected. We now turn to bias in the KLD for BBE.

Theorem 6.4 *Bias in the KLD for BBE*

A method-of-moments estimator for the KLD metric is

$$\frac{1}{2}\Big[\frac{\hat{\delta}^2}{\hat{\sigma}_T^2}+\frac{\hat{\delta}^2}{\hat{\sigma}_R^2}+\frac{\hat{\sigma}_T^2}{\hat{\sigma}_R^2}+\frac{\hat{\sigma}_R^2}{\hat{\sigma}_T^2}\Big]-1 \tag{73}$$

The expected value of expression (73) is asymptotically unbiased but is positively biased in small samples. An unbiased method-of-moments estimator is:

$$\begin{aligned}&\frac{1}{2}\Big[\frac{n-3}{n-1}\Big(\frac{\hat{\delta}^2}{\hat{\sigma}_T^2}+\frac{(n-1)\hat{\sigma}_R^2}{(n)\hat{\sigma}_T^2}-\frac{(n-1)}{(n-3)n}\Big)+\\&\frac{n-3}{n-1}\Big(\frac{\hat{\delta}^2}{\hat{\sigma}_R^2}+\frac{(n-1)\hat{\sigma}_T^2}{(n)\hat{\sigma}_R^2}-\frac{n-1}{(n-3)n}\Big)\Big]-1\end{aligned} \tag{74}$$

Proof: As method-of-moment estimators are used, taking expectations of 73) and assuming independence and using the results of Muirhead (p 24-25, 1982), it is seen that this expression reduces to:

$$\frac{n-1}{2(n-3)}\Big[\frac{\delta^2}{\sigma_T^2}+\frac{\delta^2}{\sigma_R^2}+\frac{(n+1)\sigma_T^2}{n\sigma_R^2}+\frac{(n+1)\sigma_R^2}{n\sigma_T^2}+\frac{2}{n}\Big]-1 \tag{75}$$

As sample size increases (75) becomes,

$$lim_{n\to\infty}\frac{n-1}{2(n-3)}\Big[\frac{\delta^2}{\sigma_T^2}+\frac{\delta^2}{\sigma_R^2}+\frac{(n+1)\sigma_T^2}{n\sigma_R^2}+\frac{(n+1)\sigma_R^2}{n\sigma_T^2}+\frac{2}{n}\Big]-1$$

$$=\frac{1}{2}\Big[\frac{\delta^2}{\sigma_T^2}+\frac{\delta^2}{\sigma_R^2}+\frac{\sigma_T^2}{\sigma_R^2}+\frac{\sigma_R^2}{\sigma_T^2}\Big]-1$$

which is an unbiased estimate for (67). However in small samples, the bias is

$$\frac{1}{2}\Big(\frac{n-1}{n-3}\Big)\Big[\frac{\delta^2}{\sigma_T^2}+\frac{\delta^2}{\sigma_R^2}\Big]+\frac{1}{2}\Big(\frac{(n-1)(n+1)}{n(n-3)}\Big)\Big[\frac{\sigma_T^2}{\sigma_R^2}+\frac{\sigma_R^2}{\sigma_T^2}\Big]+\frac{n-1}{n(n-3)}\geq 0$$

The proof that (74) is unbiased for (67) follows directly from these results. □□□ Thus the estimation procedure is positively biased (against sponsors) when using a 'plug-in' approach to estimation.

In similar fashion to that of the FDA BBE metric, we show that the KLD for BBE is posi-

tively biased in small samples when sample size differs between populations and asymptotically unbiased. We derive an unbiased estimator in small samples.

Theorem 6.5 *Bias in the KLD for BBE when Sample Size Differs between Populations*

When a method-of-moments estimator for the KLD of BBE is derived as in (73) where n_T and n_R are the sample size in test and reference populations, then the expected value is

$$\frac{1}{2}\Big[\frac{\nu_T}{\nu_T-2}\Big(\frac{\delta^2}{\sigma_T^2}+\frac{(n_R+1)\sigma_R^2}{(n_R)\sigma_T^2}+\frac{1}{n_T}\Big)+\frac{\nu_R}{\nu_R-2}\Big(\frac{\delta^2}{\sigma_R^2}+\frac{(n_T+1)\sigma_T^2}{(n_T)\sigma_R^2}+\frac{1}{n_R}\Big)\Big]-1 \qquad (76)$$

This expression (76) is asymptotically unbiased for (67) but is positively biased in small samples by

$$\frac{1}{n_T-3}\Big(\frac{\delta^2}{\sigma_T^2}+\Big(1+\frac{n_T}{2n_R}-\frac{1}{2n_R}\Big)\frac{\sigma_R^2}{\sigma_T^2}+\frac{1}{2}-\frac{1}{2n_T}\Big)+$$

$$\frac{1}{n_R-3}\Big(\frac{\delta^2}{\sigma_R^2}+\Big(1+\frac{n_R}{2n_T}-\frac{1}{2n_T}\Big)\frac{\sigma_T^2}{\sigma_R^2}+\frac{1}{2}-\frac{1}{2n_R}\Big)$$

An unbiased estimator of (67) is

$$\frac{1}{2}\Big[\frac{\nu_T-2}{\nu_T}\Big(\frac{\hat\delta^2}{\hat\sigma_T^2}+\frac{(n_R-1)\hat\sigma_R^2}{(n_R)\hat\sigma_T^2}-\frac{\nu_T}{(\nu_T-2)n_T}\Big)+ \qquad (77)$$

$$\frac{\nu_R-2}{\nu_R}\Big(\frac{\hat\delta^2}{\hat\sigma_R^2}+\frac{(n_T-1)\hat\sigma_T^2}{(n_T)\hat\sigma_R^2}-\frac{\nu_R}{(\nu_R-2)n_R}\Big)\Big]-1$$

Proof: Consider (73). Assuming pairwise independence and using the results of Muirhead (p 24-25, 1982), it is seen that this expression reduces to:

$$\frac{1}{2}\Big[\Big(\frac{\frac{\sigma_R^2}{n_R}+\frac{\sigma_T^2}{n_T}}{\sigma_R^2}\Big)F'_R+\Big(\frac{\frac{\sigma_R^2}{n_R}+\frac{\sigma_T^2}{n_T}}{\sigma_T^2}\Big)F'_T+\frac{\sigma_T^2}{\sigma_R^2}F_{TR}+\frac{\sigma_R^2}{\sigma_T^2}F_{RT}\Big]-1$$

where $F'_R \sim F'_{1,\nu_R}\Big(\frac{\delta^2}{\frac{\sigma_R^2}{n_R}+\frac{\sigma_T^2}{n_T}}\Big)$, $F'_T \sim F'_{1,\nu_T}\Big(\frac{\delta^2}{\frac{\sigma_R^2}{n_R}+\frac{\sigma_T^2}{n_T}}\Big)$, $F_{TR} \sim F_{\nu_T,\nu_R}$, and $F_{RT} \sim F_{\nu_R,\nu_T}$. Taking the expectation, the expression (76) is found. As sample size increases (71) becomes,

$$lim_{n_R\to\infty,n_T\to\infty}[(76)] = \frac{1}{2}\Big[\frac{\delta^2}{\sigma_T^2}+\frac{\delta^2}{\sigma_R^2}+\frac{\sigma_T^2}{\sigma_R^2}+\frac{\sigma_R^2}{\sigma_T^2}\Big]-1$$

which is asymptotically unbiased for (67).

Consider (77), which can be expressed as

$$\frac{\nu_T - 2}{2\nu_T}\Big[\Big(\frac{\frac{\sigma_R^2}{n_R} + \frac{\sigma_T^2}{n_T}}{\sigma_T^2}\Big)F_T' + \frac{(n_R - 1)\sigma_R^2}{(n_R)\sigma_T^2}F_{RT} - \frac{\nu_T}{(\nu_T - 2)n_T}\Big]$$

$$+\frac{\nu_R - 2}{2\nu_R}\Big[\Big(\frac{\frac{\sigma_R^2}{n_R} + \frac{\sigma_T^2}{n_T}}{\sigma_R^2}\Big)F_R' + \frac{(n_T - 1)\sigma_T^2}{(n_T)\sigma_R^2}F_{TR} - \frac{\nu_R}{(\nu_R - 2)n_R}\Big] - 1$$

Taking the expectation, this expression reduces to

$$\frac{1}{2}\Big[\frac{\delta^2}{\sigma_R^2} + \frac{\delta^2}{\sigma_T^2} + \frac{\sigma_T^2}{\sigma_R^2} + \frac{\sigma_R^2}{\sigma_T^2}\Big] - 1$$

□□□

6.4.4 Extensions concerning large sample properties and the inclusion of information on covariates

Now, we turn to the large sample properties of the estimates of interest. It is known (Muirhead, 1982) that the method-of-moments estimates $\hat{\delta}$, $\hat{\sigma}_T^2$, and $\hat{\sigma}_R^2$ are independent and asymptotically normally distributed unbiased estimates. Similarly, REML estimates $\hat{\delta}$, $\hat{\sigma}_T^2$, and $\hat{\sigma}_R^2$ are asymptotically normally distributed with known variances. The large sample variances for $\underline{\hat{\beta}}$ and $\hat{\boldsymbol{\Sigma}}$ are $(\boldsymbol{X}'\boldsymbol{\Sigma}^{-}\boldsymbol{X})^{-}$ and $-E[\frac{\partial^2 \boldsymbol{L}^{-1}}{\partial \boldsymbol{\Sigma}\partial \boldsymbol{\Sigma}'}]$, respectively with covariance $\mathbf{0}$ where $\boldsymbol{L}$ is the *log*-likelihood.

The arising estimates are normally distributed in the limit with variance-covariance matrix appropriate to the structure of the model. For our purposes in BBE assessment, where we estimate $\hat{\gamma} = \begin{pmatrix} \hat{\delta} \\ \hat{\sigma}_R^2 \\ \hat{\sigma}_T^2 \end{pmatrix}$, these are asymptotically normally distributed with expected value $\gamma = \begin{pmatrix} \delta \\ \sigma_R^2 \\ \sigma_T^2 \end{pmatrix}$

The asymptotic variance estimates have a symmetric variance-covariance matrix as described above, the elements of which we shall denote as

$\begin{pmatrix} l_\delta & 0 & 0 \\ 0 & l_R & 0 \\ 0 & 0 & l_T \end{pmatrix}$ Note that under this design (unlike the cross-over designs of earlier Chpaters), asymptotic estimates of the variance components σ_R^2 and σ_T^2 are independent (Searle, 1971) of each other and of the variance of $\hat{\delta}$ which we will denote l_δ.

From these findings, it is easy to derive the expected values of the relevant estimators of the FDA BBE metric and the KLD metric for BBE.

Theorem 6.6 *Asymptotic Bias and Variance of the BBE Metrics using REML Estimation*

Let

$$\hat{\nu}_{FDA} = \frac{\hat{\delta}^2 + \hat{\sigma}_T^2 - \hat{\sigma}_R^2}{\hat{\sigma}_R^2} \tag{78}$$

be an estimate for the (66) FDA metric from a REML UN model. Then, this estimate is asymptotically normally distributed, unbiased with

$$E[\hat{\nu}_{FDA}] = \frac{\delta^2 + \sigma_T^2 - \sigma_R^2}{\sigma_R^2}$$

(when $\delta \neq 0$)

Let

$$\hat{\nu}_{KLD} = \frac{1}{2}(\frac{\hat{\delta}^2}{\hat{\sigma}_T^2} + \frac{\hat{\delta}^2}{\hat{\sigma}_R^2} + \frac{\hat{\sigma}_R^2}{\hat{\sigma}_T^2} + \frac{\hat{\sigma}_T^2}{\hat{\sigma}_R^2}) - 1 \tag{79}$$

be an estimate for the (67) KLD metric from a REML UN model. Then, this estimate is asymptotically normally distributed, unbiased with

$$E[\hat{\nu}_{KLD}] = \frac{1}{2}(\frac{\delta^2}{\sigma_T^2} + \frac{\delta^2}{\sigma_R^2} + \frac{\sigma_R^2}{\sigma_T^2} + \frac{\sigma_T^2}{\sigma_R^2}) - 1$$

(when $\delta \neq 0$)

Proof: Here we apply the findings of Theorem 3.3.A of Serfling (1980) using the properties described previously of the estimates making up $\hat{\nu}_{FDA} = g(\hat{\delta}, \hat{\sigma}_T^2, \hat{\sigma}_R^2)$. The function g is obviously differentiable such that $\frac{\partial g}{\partial \hat{\delta}}\Big|_{g=\delta} = 2\delta$, $\frac{\partial g}{\partial \hat{\sigma}_T^2}\Big|_{g=\sigma_T^2} = \sigma_R^{-2}$, $\frac{\partial g}{\partial \hat{\sigma}_R^2}\Big|_{g=\sigma_R^2} = -(\frac{\delta^2}{\sigma_R^4} + \frac{\sigma_T^2}{\sigma_R^4})$.

Then by application of Theorem 3.3.A (Serfling, 1980), it is found that $g(\hat{\delta}, \hat{\sigma}_T^2, \hat{\sigma}_R^2)$ is asymptotically normally distributed with expected value $g(\delta, \sigma_T^2, \sigma_R^2)$ and variance $\underline{D}\boldsymbol{\Sigma}_l\underline{D}'$ where $\underline{D} = (2\delta, -(\frac{\delta^2}{\sigma_R^4} + \frac{\sigma_T^2}{\sigma_R^4}), \sigma_R^{-2})$ and where $\boldsymbol{\Sigma}_l$ is the Unstructured REML asymptotic variance-covariance matrix above augmented with the l_δ associated in the first row, first column, such that

$\boldsymbol{\Sigma}_l = \begin{pmatrix} l_\delta & 0 & 0 \\ 0 & l_R & 0 \\ 0 & 0 & l_T \end{pmatrix}$. The proof then follows by matrix multiplication for $\delta \neq 0$. The proof for the KLD is similar and is not reproduced here.□□□

We note here that when $\delta = 0$, the estimates are asymptotically positively biased in the same manner as the small sample estimates using method-of-moments. Simulations (to be performed later in this Chapter) will characterise the degree of bias in small sample REML estimates.

6.4.5 Inference and Example

In contrast to earlier chapters where sample sizes in excess of 20 were not unusual, for BBE studies it will be unusual to find studies with more than 8 to 12 subjects in the new region receiving a dose of drug product. Sample sizes may be substantially larger in the original region (as we shall see in the example later in this Chapter).

For the purposes of inference concerning the metrics of interest, we will utilise the nonparametric-percentile bootstrap method (Efron and Tibshirani, 1993) to construct confidence sets for the metrics of interest. From findings of previous authors as described in Shao and Tu (Chapter 3, 1995) based on the work of Bickel, Freedman, and Singh, it is known that

$$Pr[g(\hat{\gamma})_{95} > g(\gamma)] \rightarrow 0.95 \tag{80}$$

where g is the function of the estimates γ required to estimate the FDA or KLD metric of interest, where $\hat{\gamma}$ is estimated using method-of-moments, and where $g(\hat{\gamma})_{95}$ is the 95th quantile of the bootstrap distribution of the sample estimates for the metric. We know g is a differentiable function in the neighborhood of γ (see proof above), and under certain regularity conditions (discussed in Shao and Tu, 1995 and Shao et al., 2000b), it is also known that (80) holds for estimates derived from REML.

We will utilise REML estimates for the metrics of interest in BBE in this thesis so as to provide a consistent modelling approach with the later parts of this Chapter and as this is most of interest when comparing data from unbalanced populations.

An additional benefit of the use of REML is the ability to incorporate supplementary information as covariates into our estimates for the means and variances. This information is critical when for example demographic factors such as weight are of immediate concern (as we will see in the example) or when one wishes to account for the effect of dose in the findings.

In such cases, the supplementary information can be added to the REML model as a fixed effect. The resulting estimates for the means and variances are then adjusted for the effect of supplementary information (for more details see Hinkelmann and Kempthorne, Chapter 8, 1994). Our findings above with respect to the large sample properties of the arising mean and

variance estimates are adjusted for in the fixed effects space, so we find that in general, REML models will provide an efficient means of analysis in this setting, especially when supplementary information is of concern.

When not accounting for covariates, REML estimates are equivalent to method-of-moment estimates in small samples. To illustrate the concepts in involved in construction of metrics, we first consider some pharmacokinetic data from a recent submission. In the following Figure 44, AUC and Cmax (on the ln-scale) are plotted versus weight for Caucasian and Korean volunteers subjects following a single 8 mg dose.

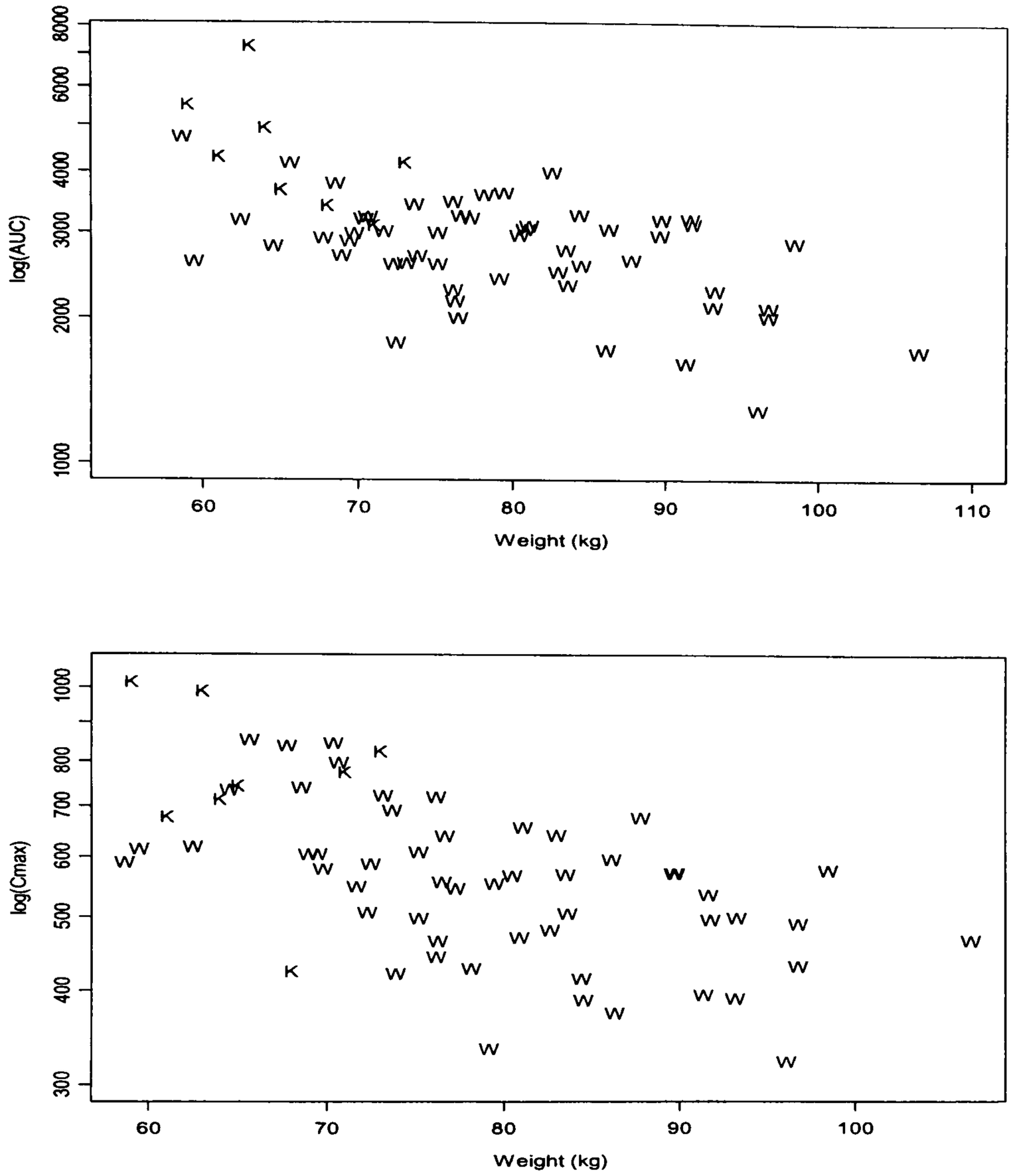


Figure 44: AUC and Cmax to Weight Response-Relationship for South Korean and Western Volunteers

Mean AUC and Cmax in South Koreans were observed to be higher than in Caucasian volunteers. Even in the most well controlled of trials, some demographic factors may differ between populations and should be accounted for in any metric constructed. In the Korean and Caucasian data set, it was found that Caucasian volunteers had greater weight than the South Korean volunteers, see the following table.

Table 25: Weight (kg) by Dose and Race

Dose	Race	N	Mean	SD	Min	Median	Maximum
8	K	8	65.5	4.84	59	64.5	73
8	W	53	79.6	10.63	58.7	78.2	106.6
K=South Korean, W=Western							

We now turn to consideration of the metrics developed thus far in this Chapter. Data at the 8 mg dose level will be examined not accounting for weight as a covariate.

Table 26: REML First and Second Moment Estimates for AUC and Cmax Data by Race

Endpoint	Race	N	$\hat{\mu}$	$\hat{\sigma}^2$
AUC	K	8	8.3785	0.0769
	W	53	7.9053	0.0632
$\hat{\delta} = 0.4732(0.2822, 0.6642)(0.3190, 0.6347)_B$				
$\frac{\hat{\sigma}_T^2}{\hat{\sigma}_R^2} = 1.2174(0.5555, 4.0363)(0.3197, 2.2020)_B$				
FDA=3.7611 (1.1537, 8.0319)				
KLD=3.2467 (2.0741, 7.7332)				
FDA'=3.3066 (0.8958, 7.3142)				
KLD'=2.4713 (1.4321, 6.0251)				
Cmax	K	8	6.6106	0.0737
	W	53	6.2990	0.0533
$\hat{\delta} = 0.3116(0.1252, 0.4979)(0.1521, 0.4551)_B$				
$\frac{\hat{\sigma}_T^2}{\hat{\sigma}_R^2} = 1.3843(0.6316, 4.5895)(0.2465, 3.4943)_B$				
FDA=2.2065 (0.7952, 4.2726)				
KLD=1.6227 (0.5830, 8.8301)				
FDA'=1.7918 (0.5035, 3.7997)				
KLD'=1.1098 (0.2289, 6.5084)				
K=South Korean, W=Caucasian				
B indicates bootstrap CI				
' Indicates Unbiased Estimator				

Here we observe that mean AUC and Cmax data (as indicated by δ) from the Koreans were significantly higher on average than that observed for Caucasians for mean AUC and Cmax. Variability appeared slightly higher in the Korean population relative to the Western population for both endpoints though we cannot conclude that variation is differentiable (see Figure 45). Examination of the density plots of bootstrapped values in Figure 45 revealed that the positive bias in the unadjusted FDA and KLD estimators is negligible relative to the unbiased estimator and variation appeared roughly similar between positively biased and unbiased estimators. As the usual estimate has been shown to be asymptotically unbiased and appeared positively biased (against sponsors) by a negligible amount, we will utilise it further in discussions and research in this thesis.

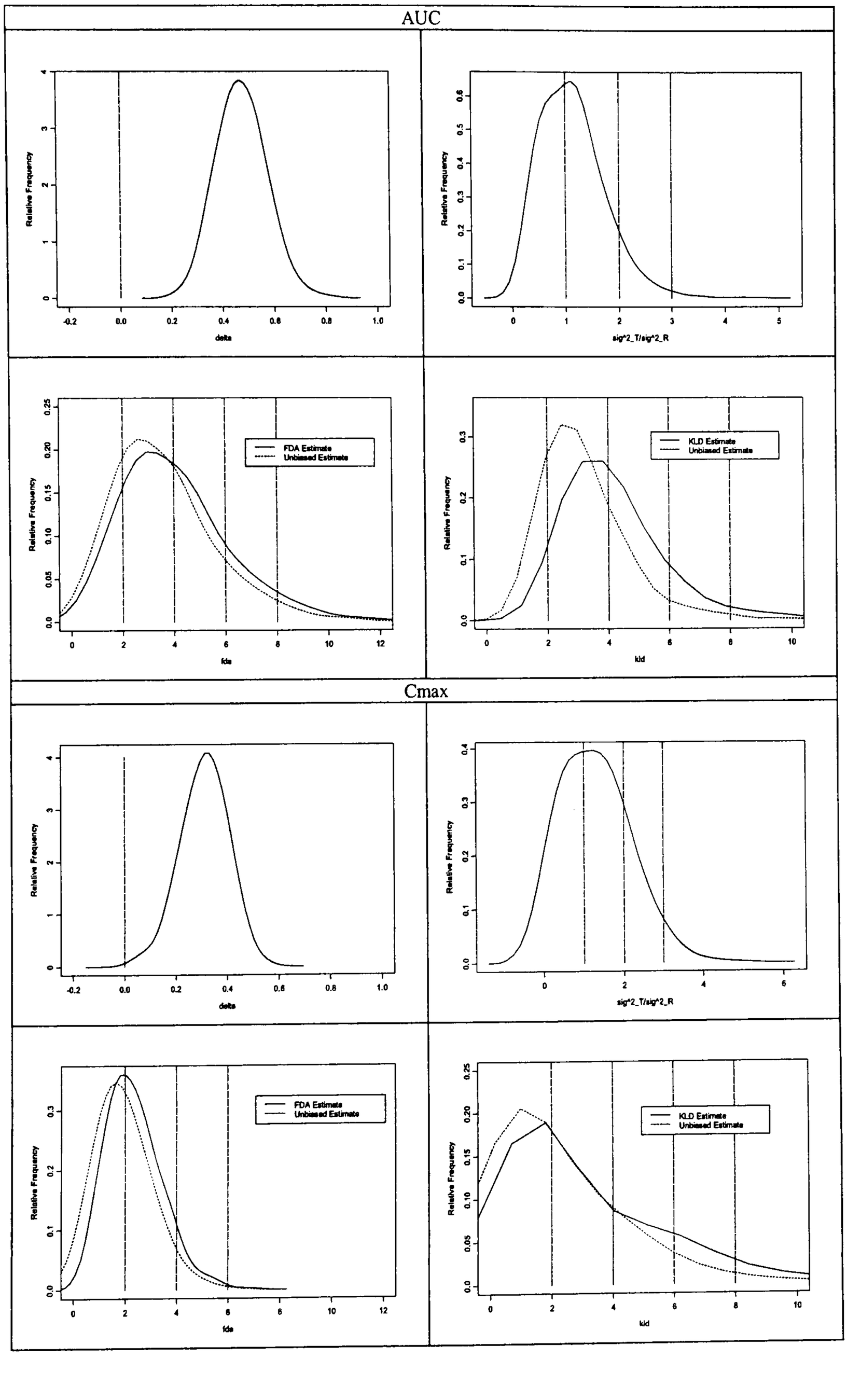


Figure 45: Density Plots of Bridging Comparisons for AUC and Cmax

We now turn to consideration of the metrics correcting for weight as a covariate.

Table 27: REML First and Second Moment Estimates for AUC and Cmax Data by Race accounting for Weight (kg) as a Covariate

Endpoint	Race	N	$\hat{\mu}$	$\hat{\sigma}^2$
AUC	K	8	8.2317	0.0612
	W	53	7.9275	0.0493
Parametric REML $\hat{\delta} = 0.3043(0.1240, 0.4845)$				
Non-Parametric REML $\hat{\delta} = 0.3039(0.1540, 0.4681)$				
$\frac{\hat{\sigma}_T^2}{\hat{\sigma}_R^2} = 1.2414(0.3338, 2.3316)$				
FDA=2.1184 (-0.0079, 6.0360)				
KLD=1.7179 (0.8517, 5.0056)				
Cmax	K	8	6.4679	0.0683
	W	53	6.3205	0.0389
Parametric REML $\hat{\delta} = 0.1474(-0.0380, 0.3328)$				
Non-Parametric REML $\hat{\delta} = 0.1446(-0.0229, 0.2960)$				
$\frac{\hat{\sigma}_T^2}{\hat{\sigma}_R^2} = 1.7524(0.3433, 3.4401)$				
FDA=1.3103 (0.1717, 3.0334)				
KLD=0.5996 (0.1860, 4.3419)				
K=South Korean, W=Caucasian				

Accounting for weight has a significant effect of the estimate of δ for Cmax but not for AUC. For both endpoints, variation appears higher in the Korean population relative to Caucasians, even when accounting for weight. Examination of the density plots of bootstrapped values in Figure 46 reveals that the densities for both AUC and Cmax are shifted to the left relative to the unadjusted estimates. Unless the acceptance levels are set very high however, it is difficult to conclude that the populations are equivalent in terms of AUC, though Cmax may be somewhat more indicative of BBE, assuming a somewhat large equivalence bound is acceptable.

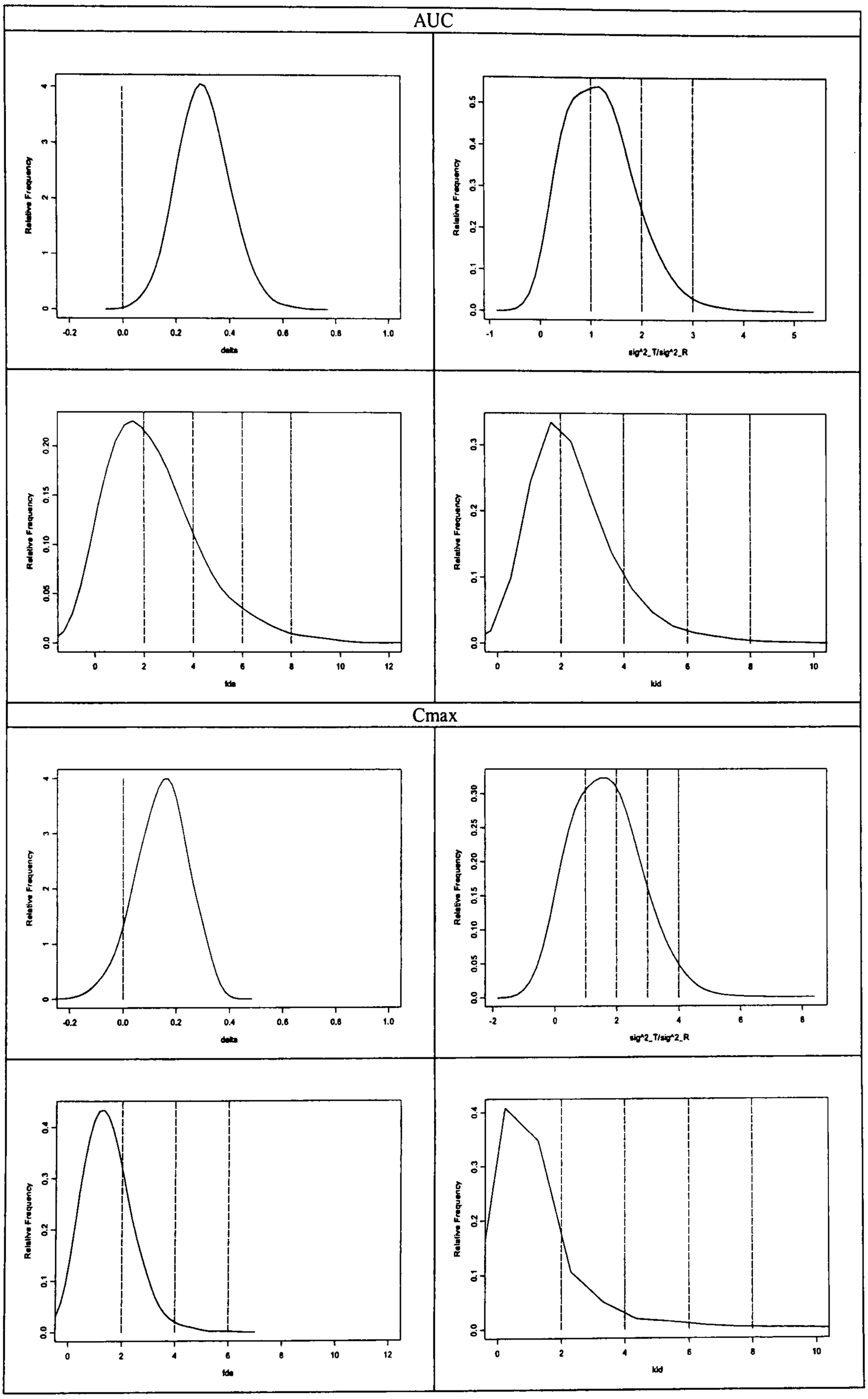


Figure 46: Density Plots of Bridging Comparisons for AUC and Cmax Accounting for Weight as a Covariate

It should be noted that when fitting linear covariates (in this example, weight), it is important (Hinkelmann and Kempthorne, 1994) to evaluate whether the assumption regarding homogeneity of slopes (Rao, 1973) is not grossly violated. In this example, $p-$values for the test evaluating this assumption (population by weight interaction) was not significant ($p > 0.05$) and was not indicative of heterogeneity of slopes indicating that comparisons between populations adjusted for these factors are statistically valid (Hinkelmann and Kempthorne, 1994). It should be noted however that this evaluation is relative to the uncertainty observed in the study and that the study was not designed to assess this factor.

6.4.6 Simulation

In this section, we will use simulation to:

1. Assess whether REML estimation results in negligible bias in small samples for δ, σ_T^2/σ_R^2, and the FDA and KLD metrics.

2. Compare bounds from the traditional approach to inference versus the nonparametric percentile bootstrap method for δ and σ_T^2/σ_R^2.

Two group (T or R), single PK outcome designs were simulated using $SAS^{©}$ using the parameter space described in the below table. Simulations were conducted for sample sizes of $(n_T, n_R) = (8, 24)$ and $(n_T, n_R) = (16, 48)$ in accord with the practicalities of sample size discussed earlier in this section. Each simulation study was composed of 500 runs. Nonparametric percentile bootstrapping was conducted with 1000 samples drawn for each run of the simulated data sets. The parameter space studied is defined in the below table.

Table 28: True Values used in Simulation Experiments 1 through 36 (500 runs per simulation)

Sim	δ	$\frac{\sigma_T^2}{\sigma_R^2}$	σ_R	(n_T, n_R)
1	0	0.5	0.1	(8,24)
2	ln 1.5	0.5	0.1	(8,24)
3	ln 2	0.5	0.1	(8,24)
4	0	1	0.1	(8,24)
5	ln 1.5	1	0.1	(8,24)
6	ln 2	1	0.1	(8,24)
7	0	3	0.1	(8,24)
8	ln 1.5	3	0.1	(8,24)
9	ln 2	3	0.1	(8,24)
10	0	0.5	0.8	(8,24)
11	ln 1.5	0.5	0.8	(8,24)
12	ln 2	0.5	0.8	(8,24)
13	0	1	0.8	(8,24)

Table 28: True Values used in Simulation Experiments 1 through 36 (500 runs per simulation)

Sim	δ	$\frac{\sigma_T^2}{\sigma_R^2}$	σ_R	(n_T, n_R)
14	ln 1.5	1	0.8	(8,24)
15	ln 2	1	0.8	(8,24)
16	0	3	0.8	(8,24)
17	ln 1.5	3	0.8	(8,24)
18	ln 2	3	0.8	(8,24)
19	0	0.5	0.1	(16,48)
20	ln 1.5	0.5	0.1	(16,48)
21	ln 2	0.5	0.1	(16,48)
22	0	1	0.1	(16,48)
23	ln 1.5	1	0.1	(16,48)
24	ln 2	1	0.1	(16,48)
25	0	3	0.1	(16,48)
26	ln 1.5	3	0.1	(16,48)
27	ln 2	3	0.1	(16,48)
28	0	0.5	0.8	(16,48)
29	ln 1.5	0.5	0.8	(16,48)
30	ln 2	0.5	0.8	(16,48)
31	0	1	0.8	(16,48)
32	ln 1.5	1	0.8	(16,48)
33	ln 2	1	0.8	(16,48)
34	0	3	0.8	(16,48)
35	ln 1.5	3	0.8	(16,48)
36	ln 2	3	0.8	(16,48)

These simulations will also be used to characterise the coverage probability rates in small samples for δ, σ_T^2/σ_R^2, and the FDA and KLD metrics.

The second set of simulations were performed in order to investigate the Type II error rates of the BBE procedures for σ_T^2/σ_R^2, and the FDA and KLD metrics for a range of sample sizes. Sample size and power for equivalence testing of δ using confidence intervals has been described previously in Diletti et al. (1991; see also Hauschke, 2002 for a bibliography) and will not be further explored in this thesis.

Table 29: True Values used in Simulation Experiments 37 through 54 (500 runs per simulation, 1000 bootstraps per run) where $n_R = 60$

Sim	δ	σ_T	σ_R	n_T
37	ln 1	0.1	0.1	16
38	ln 1	0.1	0.1	20
39	ln 1	0.1	0.1	30
40	ln 1	0.1	0.1	40
41	ln 1	0.1	0.1	50
42	ln 1	0.1	0.1	60
43	ln 1	0.3	0.3	16
44	ln 1	0.3	0.3	20
45	ln 1	0.3	0.3	30
46	ln 1	0.3	0.3	40
47	ln 1	0.3	0.3	50
48	ln 1	0.3	0.3	60
49	ln 1	0.5	0.5	16
50	ln 1	0.5	0.5	20
51	ln 1	0.5	0.5	30
52	ln 1	0.5	0.5	40
53	ln 1	0.5	0.5	50
54	ln 1	0.5	0.5	60

Simulations were performed using the $SAS©$ procedure 'rannor' in a manner appropriate to a parallel group design and were performed using $SAS©$ version 8.1 running under UNIX. The programme used to perform the simulations and bootstrapping may be found in the appendix. $SAS©$-based REML mixed modelling procedures and method-of-moments modelling were conducted in accordance with the descriptions in this Chapter.

We now turn to the findings from the simulation exercise. We first address the issues of bias in estimates of δ, the ratio of variances, and the BBE metrics. This will be followed by discussion of findings relating to Type I and II error in BBE assessment.

6.4.7 Estimation and Precision of δ and $\frac{\sigma_T^2}{\sigma_R^2}$

Findings from simulation studies 1-36 may be found in Table 98. Based on the simulations, in the smallest of studies ($n_T = 8$, $n_R = 16$) one should expect a small degree of bias in the estimate for δ; however, this bias is minor (around 5% in the simulations studied) and would not be expected to impact inference. If sample size is increased (to $n_T = 16$, $n_R = 48$), the estimated bias becomes negligible.

In the smallest of studies ($n_T = 8$, $n_R = 16$) one should expect a small degree of positive bias in the estimate for $\frac{\sigma_T^2}{\sigma_R^2}$; however, this bias is minor and decreases with increasing sample size. The magnitude of the bias would not be expected to impact inference.

The mean (SE) and median (5th and 95th quantiles) of the lower and upper 90% bounds for δ and $\frac{\sigma_T^2}{\sigma_R^2}$ derived using parametric methods and the nonparametric percentile bootstrap method may be found in Table 99.

For δ, the nonparametric bootstrap lower bound appears shifted to the right relative to the traditional bound (computed based on a t-interval with Kenward-Roger (1997) degrees of freedom) based on inspection of the mean and median findings. The upper bound for the nonparametric bootstrap similarly appears shifted left relative to the traditional upper bound. However, the spread in the estimates appears similar between statistical methods (as evidenced by the SE and quantile findings). This shift appears to decrease with increased sample size (see simulations 19-36) suggesting that the findings for the bootstrap are associated with asymptotic nature of the bootstrap's probability coverage, and this finding will be investigated in the subsequent assessments of this chapter.

For $\frac{\sigma_T^2}{\sigma_R^2}$, the nonparametric bootstrap lower and upper bounds appear shifted to the left relative to the lower and upper bounds derived using the traditional F-statistics. Additionally, the traditional bounds appear more variable (as evidenced by the SE and quantiles) relative to the bootstrap confidence limits. As with the nonparametric percentile bootstrap bounds for δ, this suggests that the asymptotic nature of probability coverage for the bootstrap may impact the results in small samples. This will be investigated subsequently in this chapter and an algorithm for calibration will be developed. This shift appears to decrease with increased sample size (see simulations 19-36).

6.4.8 Estimation and Precision of the FDA and KLD metrics for BBE

Findings from simulation studies 1-36 may be found in Table 98. Based on the simulations, in the smallest of studies ($n_T = 8$, $n_R = 16$) one should expect a potentially large degree of bias in the estimate for the FDA and KLD metrics. In the smallest of studies, this bias can be quite large and may impact inference. However, the bias is positive (against sponsors) making it 'harder' to demonstrate BBE. If sample size is increased by a small amount (to $n_T = 16$, $n_R = 48$), the estimated bias becomes negligible and would not be expected to impact inference in the majority of situations. This raises the possibility however that coverage probability may be impacted, and this will be investigated in the following section of this chapter.

6.4.9 Coverage Probability Rates

The true values for δ, $\frac{\sigma_T^2}{\sigma_R^2}$, and the FDA and KLD metrics were assessed relative to the traditional and nonparametric percentile bootstrap confidence bounds for simulations 1-36 in Table 30 below. The traditional 90% t-interval for δ was derived using the Kenward-Roger (1997) degrees of freedom and 5th and 95th quantiles of the nonparametric bootstraps were selected as the corresponding bounds (Efron and Tibshirani, 1993). The traditional 90% F-interval for $\frac{\sigma_T^2}{\sigma_R^2}$ was derived as described previously in this section, and 5th and 95th quantiles of the nonparametric bootstraps were selected as the corresponding bounds. For the FDA and KLD metrics, the 95th quantile of the nonparametric bootstraps were utilised as the upper bound for testing purposes in accordance with previous research in Chapters 3 and 4.

Table 30: Coverage Probability Rates for True Values of δ, $\frac{\sigma_T^2}{\sigma_R^2}$, and the FDA and KLD metrics in Simulation Experiments 1 through 36 (500 runs per simulation with 1000 Bootstraps per run)

Sim	δ T	$\frac{\sigma_T^2}{\sigma_R^2}$ T	δ NP	$\frac{\sigma_T^2}{\sigma_R^2}$ NP	FDA NP	KLD NP
1	89	90.2	86	81.8	97.2	100
2	89	90.2	86	81.8	97.6	100
3	89	90.2	86	81.8	98	99.8
4	89.4	90.2	84.8	81.8	95.4	100
5	89.4	90.2	84.8	81.8	96.8	100
6	89.4	90.2	84.8	81.8	97.6	100
7	89.8	90.2	84	81.8	93	100
8	89.8	90.2	84	81.8	94.6	99.6
9	89.8	90.2	84	81.8	96.8	99.8
10	89	90.2	86	81.8	97.2	100
11	89	90.2	86	81.8	94.6	100
12	89	90.2	86	81.8	92.6	100
13	89.4	90.2	84.8	81.8	95.4	100
14	89.4	90.2	84.8	81.8	93	100
15	89.4	90.2	84.8	81.8	91.6	100
16	89.8	90.2	84	81.8	93	100
17	89.8	90.2	84	81.8	91.2	100
18	89.8	90.2	84	81.8	91.4	100
19	89.2	89.6	89	82.6	96	100
20	89.2	89.6	89	82.6	95.2	99.8
21	89.2	89.6	89	82.6	96.2	99.6
22	87.6	89.6	86	82.6	92.8	100
23	87.6	89.6	86	82.6	95.4	99.6
24	87.6	89.6	86	82.6	95.8	100
25	87.8	89.6	85.2	82.6	91.8	98.6
26	87.8	89.6	85.2	82.6	93.4	98
27	87.8	89.6	85.2	82.6	95.4	99.4
28	89.2	89.6	89	82.6	96	100
29	89.2	89.6	89	82.6	93	99.4
30	89.2	89.6	89	82.6	93.8	98.8
31	87.6	89.6	86	82.6	92.8	100
32	87.6	89.6	86	82.6	90.8	100
33	87.6	89.6	86	82.6	91.8	99.6
34	87.8	89.6	85.2	82.6	91.8	98.6
35	87.8	89.6	85.2	82.6	89.8	95.8
36	87.8	89.6	85.2	82.6	90.8	95.8

Trad: Traditional t- or F- confidence bounds
NP: Nonparametric Percentile Bootstrap confidence bounds

As expected, the traditional t-interval (for δ) and F-interval (for $\frac{\sigma_T^2}{\sigma_R^2}$) confidence bounds resulted in coverage probability rates of approximately 90% consistent with their properties of first-order correctness (Efron and Tibshirani, Ch. 18, 1993). Coverage probability rates of the nonparametric percentile method were lower than the traditional approach as expected in keeping with the small sample size, asymptotic findings with regard to coverage probability rates, and in the knowledge that the nonparametric bootstrap intervals are themselves only first-order correct; however, they did not appear to fall below 80% and should serve adequately for descriptive purposes. For δ, increasing sample size caused the nonparametric coverage to approach the traditional rate within 2-3% and should prove adequate for formal statistical testing (which is not unexpected given the asymptotic basis for the bootstrap). Increasing sample size also appeared to cause the coverage probability rate for the nonparametric percentile bootstrap to approach the traditional rate for $\frac{\sigma_T^2}{\sigma_R^2}$; however, in the sample sizes studied, the coverage probabilities for the nonparametric bootstrap were still 6-8% lower than the gold-standard F-intervals. This lower than expected rate was not unexpected when the asymptotic normality basis for the nonparametric bootstrap procedure is taken into account.

To correct this deficit in the nonparametric percentile bootstrap coverage probability rate, we advise that a minimum of 12-16 subjects in the test population be exposed to drug to ensure that the statistics used provide an appropriate coverage rate for δ; furthermore, if estimation and-or testing involving the variances is warranted, we advise that the nonparametric percentile interval be calibrated (Efron and Tibshirani, Ch. 18, 1993). In Efron and Tibshirani (1993), a data driven bootstrap-based algorithm is utilised; however, it is unlikely such an approach would be acceptable to regulatory agencies as the calibration is based upon the data collected. We recommended an alternative algorithm that is simulation-based to determine (a priori) the appropriate quantiles to select from the nonparametric bootstraps to provide a confidence interval of appropriate 90% coverage:

Algorithm 6.1

1. Simulate X data sets (x_1, x_2,,, x_X) of sample size consistent with the study design planned on the basis of the known value of θ (where θ is the unknown parameter of interest).

2. For each simulated data set x_X, generate B bootstrap samples (x_X^{*1},,,, x_X^{*B}).

3. Derive λ and $1-\lambda$ level confidence points for $\hat{\theta}_\lambda(b)$ and $\hat{\theta}_{1-\lambda}(b)$ for a grid of values of λ for each data set x_X.

4. For each λ and $1-\lambda$, compute $p(\lambda) = \#\{\theta \leq \hat{\theta}_\lambda(b)\}/X$ and $p(1-\lambda) = \#\{\theta \leq \hat{\theta}_{1-\lambda}(b)\}/X$.

5. Choose the values of λ_X and $1-\lambda_X$ from the grid of λ such that $p(\lambda_X) = 0.05$ and $p(1-\lambda_X) = 0.95$.

One can easily then confirm via simulation that these levels of λ_X and $1-\lambda_X$ will provide a 90% coverage probability rate. Consider the example provided previously in this chapter. If we perform a simulation (500 runs with 1000 bootstraps per run) using the estimated parameters in this example, under Algorithm 6.1, it is found that the 13th and 100th quantiles should be used to determine a 90% nonparametric bootstrap confidence interval for $\frac{\sigma_T^2}{\sigma_R^2}$, and the resulting interval based on the data is found to be $(0.48, 4.61)$ for AUC closely approximating the bounds generated for the F-interval which we have observed to provide approximately 90% coverage in this type of study when no covariates are present. Note however, that in our example simulation the 100th percentile is selected by the algorithm suggesting that not enough bootstraps have been taken to fully characterise the distribution of responses in the tail. In this situation, we recommend that more bootstraps be taken in practice (eg. a minimum of 2000). Computationally this can be challenging, but modern computing power should provide sufficient resource to do this in this relatively simple situation.

For the FDA and KLD metrics, conservative coverage probabilities were sometimes observed consistent with previous findings in Chapter 5. Coverage probabilities for the FDA metric appears provide at least a 90% coverage rate. The KLD coverage probability rate, on the other hand, appears conservative and maintains at least 95% coverage. These findings appear to be related to the inclusion of both $\frac{\sigma_T^2}{\sigma_R^2}$ and $\frac{\sigma_R^2}{\sigma_T^2}$ in the KLD; these two terms are so ill characterised in samples this small that the upper bound of KLD is inflated in recognition of this combination. As we shall see in the next section, sample sizes adequate to provide sufficient power for goalposts for equivalence testing are quite large and will likely not be acceptable for use of such testing of composite metrics in this setting when it is remembered that a primary purpose of these BBE studies is to confirm that it is safe to study large numbers of subjects and patients in the new region in further drug development.

If formal inference is being drawn, an approach to statistical analysis should be used to ensure coverage probability is conserved. Traditional methods appear to provide an appropriate level of coverage in this setting for δ and $\frac{\sigma_T^2}{\sigma_R^2}$, but such parametric methods are only appropriate for $\frac{\sigma_T^2}{\sigma_R^2}$ assessment when no covariates are utilised in the restricted maximum likelihood model. When covariates are included, calibrated nonparametric percentile bootstrap intervals provide adequate coverage and can easily be constructed using simulation-based methods for these and alternative metrics.

6.4.10 Type II Error in BBE Testing of σ_T^2/σ_R^2 and the FDA and KLD metrics

The selection of goalposts in this report is consistent with the FDA's approach to goalpost setting for PBE and implies 50% increase or acceptance range of (0.67, 1.5) for σ_T^2/σ_R^2. Simple inspection of an F-table reveals that in these type of studies with the sample sizes available, testing for equivalence based on this range is not feasible. Even for an $n_R = n_T = 61$ the 95th quantile of the $F_{60,60}$-distribution is 1.53. For the purposes of this report, however, we will utilise the range (0.5, 2) to provide a gross range within which it may be desirable to assess equivalence of variances.

If the original ideas proposed by FDA for goalpost setting are followed (FDA Guidance, 1997), then the BBE acceptance bound for the FDA metric would be 1.74. As can see from Figure 42, this allows quite wide trade-offs in variances as δ is usually small compared to σ_R^2. We will utilise this bound and a more conservative bound (1, selected based on inspection of Figure 42) in this assessment. For KLD we will utilise a bounds of 1 (conservative) and 1.5 as this seems reasonable based on inspection on Figure 43.

The findings of the exercise may be found in the following Table 31:

Table 31: Failure Rates (%) for Simulation Experiments 37 through 54 (500 runs per simulation, 1000 bootstraps per run) where $n_R = 60$ for BBE Metrics and Acceptance Criteria

Sim	$\frac{\sigma_T^2}{\sigma_R^2}$ (0.5,2)	FDA 1	FDA 1.74	KLD 1	KLD 1.5
37	86.8	47.2	17.6	46.0	18.8
38	79.6	41.4	12.4	24.6	9.0
39	56.0	34.6	8.2	9.2	0.8
40	41.2	27.4	4.6	2.4	0.2
41	37.6	25.4	2.8	1.0	0
42	34.8	23.2	2.2	0.4	0
43	86.8	47.2	17.6	46.0	18.8

Table 31: Failure Rates (%) for Simulation Experiments 37 through 54 (500 runs per simulation, 1000 bootstraps per run) where $n_R = 60$ for BBE Metrics and Acceptance Criteria

Sim	$\frac{\sigma_T^2}{\sigma_R^2}$ (0.5,2)	FDA 1	FDA 1.74	KLD 1	KLD 1.5
44	79.6	41.4	12.4	24.6	9.0
45	56.0	34.6	8.2	9.2	0.8
46	41.2	27.4	4.6	2.4	0.2
47	37.6	25.4	2.8	1.0	0
48	34.8	23.2	2.2	0.4	0
49	86.8	47.2	17.6	46.0	18.8
50	79.6	41.4	12.4	24.6	9.0
51	56.0	34.6	8.2	9.2	0.8
52	41.2	27.4	4.6	2.4	0.2
53	37.6	25.4	2.8	1.0	0
54	34.8	23.2	2.2	0.4	0

From these simulations we conclude that power to demonstrate equivalence using the population variances is very low in such studies, unless a large acceptance interval or value is used. Alternatively, one may increase the sample size to provide adequate power, but this is likely to be less than acceptable in early phase studies.

6.5 Discussion

In ICH-E5 comparisons of pharmacokinetics between populations, it is important to recognise that the presence of confounding factors relating to intrinsic differences in populations will be involved in inference. Traditional methods for the comparison of pharmacokinetics involving the population means are sufficient if this is not the case, but in general, more sophisticated statistical methods should be utilised to account for these intrinsic factors. Clinical development planning, study design, and the role of pharmacokinetics in this setting are described in this chapter.

Restricted maximum likelihood based modelling using the nonparametric percentile bootstrap for testing, appears adequate for comparison of the means between populations. Power to demonstrate a pre-specified equivalence acceptance region for the difference in population means was previously discussed in Diletti et al. (1991) and appears feasible for some drug products. Estimates for these and other metrics from REML are nearly unbiased in small samples and are known to be asymptotically unbiased in distribution. The nonparametric percentile bootstrap appears to provide adequate probability coverage and should serve for most purposes when covariate information is included in a REML model.

Extensions to this approach to estimation and-or testing on the variances however are more challenging. While testing of the variances may be important in a clinical sense, it is extremely unlikely such studies will be able to differentiate variance statistically between populations. In such cases, one should consider interpretation of the data based on the plausible range of values (Hauck and Anderson, 1986). Estimates for the ratio of population variances and other more sophisticated metrics involving both means and variances as developed in this report are nearly unbiased in small samples and are asymptotically unbiased, but the nonparametric bootstrap provides slightly low probability coverage in this situation relative to the desired rate and should be calibrated using a simulation based method prior to data analysis. Moreover, the power to

demonstrate equivalence based on the goalposts considered in this chapter is either not feasible or is likely to involve large sample sizes.

Pharmacokinetic equivalence is not a prerequisite for successful bridging using pharmacokinetics. Interpretation of pharmacokinetic changes in light of the dose (concentration) - response relationship is pivotal. Shifts in the mean of pharmacokinetic parameters measured will often be observed, and it will be important to consider their clinical, rather than statistical relevance. Goalpost setting, as applied in the context of the analysis of bioequivalence, precludes the drawing of conclusions based on broader issues related to the drug itself. It will therefore be of limited interest as drugs will be considered on a case-by-case basis in light of all relevant issues. Furthermore, the subject numbers required for studies providing goalpost evidence are large.

The initial simulation studies summarised in this report consider bias in the statistics as one of the primary outcomes of interest in line with the use of this finding in bioequivalence related simulations. Future simulations will consider alternative approaches to consideration of the simulation findings. In particular mean squared error and the probability that the absolute value of the estimate falls beyond some constant ϵ will be considered in future research. It is also worthy of mention that the simulations assume data arises from a normal distribution. While this assumption is reasonable for pharmacokinetic data (Westlake, 1986), for other data this may not be the case and may impact inference.

Extension of these techniques to other situations involving multiple doses and pharmacodynamic or clinical data should prove possible and will constitute the topics of future research. Of particular interest, application of the Kolmogorov-Smirnov test for two independent samples (Sheskin, 2000) and the techniques described by O'Brien (1988) seem applicable in this setting and will be considered.

7 Conclusions

This thesis explores the statistical science of what is involved in bioequivalence testing, of how one goes about designing, powering, and analysing studies, and where such techniques are implemented in drug development. The thesis also describes why such studies and their analyses are performed and who is involved as a stakeholder in the outcome. We will briefly summarise the findings of this thesis in this chapter and consider topics for future research.

The science of biopharmaceutical statistics traditionally has focused on differentiating between products (or placebo) to provide new and enhanced treatments for the public's benefit (Senn, 1997). However, this is generally expensive and time-consuming (DiMasi, 2001) and over time steps have been taking to reduce costs and to increase supply of pharmaceutical products while maintaining the potential for innovation. One such example pertains to bioequivalence.

An history of bioequivalence may be found in Chapter 1. To summarise, in the 1970s and 1980s, advances in chemical and biopharmaceutical engineering made it relatively feasible to create improved formulations of drug products (to enhance efficacy while maintaining or improving safety profile) and to create inexpensive copies of pharmaceutical products, which could presumably be marketed once patent protection on the innovator product was exhausted. Biopharmaceutical statistics then was required to assess how one would go about quantitatively defining how the two products could be determined to be equivalent.

To call something equivalent implies a context or criteria for the determination. The US Food and Drug Administration were directed to create such a context by the US government in 1984. There are several stakeholders in determining such a criteria:

* Regulatory and public-health considerations: The approach used must protect public health (in that the risk of false positive market access must be controlled at a pre-determined rate).

* Statistical considerations: the approach should be quantifiable, accurate, precise, well understood, and should be transparent in interpretation.

* Sponsor considerations: Using a well-designed, controlled, and reasonably sized study (or set of studies) the sponsor should be able to show the criteria have been met with a quantified chance of success.

Various approaches to the problem of bioequivalence were considered in the 1980 through early 1990s and were discarded as they failed to address one or more of the above considera-

tions. This debate culminated in 1987 when Schuirman's two one-sided testing method for a regulatory set goalpost of 20% was introduced using pharmacokinetic measures AUC and Cmax as surrogate markers for efficacy and safety by the FDA. The design of choice was determined to be a randomised two-period cross-over in normal healthy volunteers to isolate and quantify any differences in formulation, and regulatory risk was set at 5% per test. The design and analysis of cross-over studies had been extensively developed by this time (Jones and Kenward, 1989; Senn, 1993; Senn, 2002), and statistical considerations in power and sample size were described in Diletti et al. (1991).

This approach was formalised in the 1992 FDA Guidance and applied to both pre- and post-marketing approvals for changes in formulation. Average bioequivalence quickly became an international standard with most nations utilising the FDA's 1992 guidance or slight modifications to the approach. To date, products which have utilised this approach have not been observed to have marketplace failures in terms of their safety and efficacy profiles (see Barrett et al., 2000 for more details). Average bioequivalence testing of $\delta = \mu_T - \mu_R$ thus has been established de facto as a surrogate marker for public safety based upon primarily upon observation, consistency of knowledge, and replication of findings of the application of the FDA Guidance (1992) and less upon quantified, scientific assessment of biological plausibility and strength of association.

Average bioequivalence did however have the potential for issues in implementation with regard to the three considerations above. One potential difficulty was regulatory in nature. The approach was concerned with testing only the formulations means and did not contain any explicit criteria pertaining to individual subjects, and it was felt that the inclusion of criteria relating to variation might address such points. Another potential area of difficulty involved both regulatory and sponsor considerations. The regulatory limits of 20% were also questioned as they might be too large for low variability products with a narrow therapeutic index, and the 20% acceptance limits created a practical difficulty for sponsors in that sample size became very large to have a high probability of success for high variability products.

FDA addressed this second issue presented by low variability drugs by tightening the range in some instances (eg. for vaccines), and it was known alternative designs (Vonesh and Chinchilli, 1997) and mixed modelling approaches (Satterthwaite, 1941; Kenward and Rogers, 1997) could

in theory be used to demonstrate average bioequivalence to address sponsor's considerations for highly variable drug products, though the statistical and regulatory considerations of such an approach were not precisely defined. FDA opened the discussion on the resolution of these issues with the publishing of the 1997 preliminary draft guidance and significant international debate followed (for examples see Chapter 1).

Sponsor, Statistical, and Regulatory considerations in meeting the average bioequivalence criteria with a replicate design study were considered in Chapter 2 and 5. Statistically, it was determined that a constrained restricted maximum likelihood approach could be used in small samples to assess average bioequivalence in a well-controlled and replicate design study and that the Satterthwaite (1941) or Kenward-Roger approximation (1997) protected the regulatory risk of Type I error at less than 5% in average bioequivalence testing. The choice of REML variance-covariance structure was observed to have the potential to impact inference (Chapter 2). Utilisation of REML techniques to provide unbiased variance estimates in small samples from replicate designs (with or without missing data) were determined, and small and large sample estimation and Type I and II error characteristics for average bioequivalence were precisely defined in Chapters 2 and 5 of this thesis when using a replicate design. Average bioequivalence assessment and interpretation is thus readily transparent through the use of 90% confidence intervals and the practical implementation of such techniques using this higher order design is now well understood.

In 1997, FDA also reconsidered criteria to address the explicit inclusion of criteria to include individual subject responses. This was addressed in the individual bioequivalence criteria by the inclusion of variance components for subject-by-formulation interaction (Ekbohm and Melander, 1989) and within-subject variances (Chinchilli and Esinhart, 1996). International debate on the merits of this proposal was extensive (see Chapter 1 for details). Regulatory considerations were in part addressed through the use of a replicate cross-over design for this assessment and using a one-sided test 5% for the criterion, though the setting of acceptance criteria in this setting were not well defined. Statistical and sponsor considerations, however, were not well-defined in the proposal and were the subject of extensive, vociferous debate.

The first of many statistical issues concerned the estimation method to be used in the assessment of IBE. It was known that a certain constrained REML model (CSH option) was potentially

inappropriate for this assessment, and the use of the nonparametric percentile bootstrap was not well understood in this confirmatory setting.

These topics were not well studied and were quickly (and quietly) discarded by the FDA in subsequent guidances (1999a-b, 2000, 2001) as alternatives were proposed based on method-of-moments estimation with inference to be based on confidence limits generated using a Cornish-Fisher approximation (Hyslop et al., 2000). Additionally, PBE was not studied save for isolated reports. Techniques were developed and studied for its testing in this thesis.

This thesis made an extensive study of REML and method-of-moments estimation in replicate designs and showed, based on retrospective analysis and simulation, that when the variance components and differences between means are unbiased, the 'plug-in' estimates of the IBE and PBE criteria were biased (Chapters 3 and 4) by a negligible amount in small samples and were asymptotically unbiased and normal in distribution. However, significant bias in the estimates of the differences in means and in the variance components may be introduced by the presence of certain patterns of missing data when method-of-moments estimation is used and by the placement of constraints on the parameter space when using REML estimation (Chapter 5).

REML models were proposed to deal with the presence of these factors (Chapter 2), and the impact on Type I and II error was studied in this setting using simulation (Chapter 5). It was determined that the Hyslop et al. (2000) procedure was appropriate for most situations and, where it was not as determined by the patterns of missing data, REML methods using asymptotic tests or alternatively calibrated nonparametric-percentile bootstrapping (under certain circumstances) were shown to provide alternative procedures protecting sponsor and regulatory considerations and maintaining the statistical validity of the findings.

However, the use of the IBE and PBE criteria were not found to be transparent and the validity of the inclusion of variance components making up the IBE and PBE criteria was questionable. The findings of Chapter 2-3 indicated that the precision of the subject-by-formulation variance (σ_D^2) involved in IBE testing was extremely poor, and found estimates were frequently negative or null unless constraints were placed on their estimation. The imposition of such constraints however introduces bias. These findings were enhanced using simulation in Chapter 5, and this thesis concludes that σ_D^2 is a poor surrogate for individual switchability based upon lack of observations indicative of temporal observation of a problem (Chapter 1), consistency of

knowledge, replication of findings, and strength of association (Chapters 2, 3, and 5). Biological plausibility was addressed by other authors (Chen et al., 2000a-b; Hauck et al., 2000) and was not considered in this thesis.

Between-subject and within-subject variance estimates are similarly poorly characterised (Chapter 3 and 4) and their use in the composite criteria led to results which were likely not what regulators intended. For example, PBE can be demonstrated for highly variable drugs with changes in mean exposure between formulations of at least 40% (Chapter 4) with reasonable probability of success (Chapter 5). IBE allows for the potential for public health risks to be created in generic-to-generic switching (Chapter 5) which are not present when using the ABE criteria (Anderson and Hauck, 1996).

We recommend that these IBE and PBE criteria not be used as proposed. The potential for public risk was recognised in part based upon the work done for this thesis, and IBE and PBE were removed from FDA guidance in 2002. The approach used to define of the acceptance criteria for IBE and PBE (Chapter 3 and 4) are likely a contributing factor to this finding, and if IBE and PBE are ever re-interrogated as potential approaches for market access, further research in these definitions will be worthy of attention.

Assuming one does condition on the value observed for reference product variation using an appropriate test, impact of the potential dependence on expectation and variance is recognisable. This however is not the framework taken by the regulatory agency which is the basis for the thesis. Such however consititutes an interesting area for future research.

A related problem is the question of equivalence testing for pharmacokinetic data between populations. This approach impacts the regulatory acceptance of foreign pharmacokinetic data, in general to provide supporting information for larger studies, but for some few drug products, comparison of such data may constitute a basis for approval to market. Parallel-group designs will be used for this purpose, and this thesis (Chapter 6) describes the drug development plan along with study design issues which should be considered to address the Regulatory, Statistical, and Sponsor considerations which were described earlier in this chapter. A model-based, calibrated, nonparametric percentile bootstrap approach was developed in this thesis to allow for the inclusion of supporting information (i.e. covariates) and shown to provide appropriate estimates and coverage probability rates for the metrics of interest in bioequivalence testing

in this setting. Power and sample size requirements for bioequivalence testing were developed (building on the work of Diletti et al., 1991). A simulation based environment and an algorithm for future research in this area was created (Chapter 6).

On a practical level, it is unlikely that pharmacokinetic studies will serve as sufficient for market access except in rare circumstances (an example where it was sufficient is provided in Chapter 6). Power and sample size for the assessment of equivalence is larger than would in general be used for such small, well-controlled pharmacokinetic studies, and it is likely that only the difference in population means will be characterised to an appropriate degree of precision in this setting. Thus if more information is needed, we recommend a larger trial involving dose response in patients should be performed. Statistical techniques for this type of study will be considered in future research.

This thesis advances the understanding of regulatory, sponsor, and statistical issues in average, individual, and population bioequivalence testing. The findings represent the first comprehensive comparison of estimation methods to determine accuracy and precision in replicate designs following on from the work of Jones and Kenward (1989), Senn (1993), Vonesh and Chinchilli (1997), and Senn (2002). A simulation based environment is provided to answer questions arising for practical research. This thesis also contains the first comprehensive comparative study of Type I and II error rates in Cornish-Fisher, Nonparametric percentile bootstrap, and asymptotic testing for individual and population bioequivalence, and additionally thoroughly explores the practical implications of missing data. Finally, the potential for public health risk is assessed for techniques proposed by FDA using simulation and found to be quite substantial and of concern sufficient to result in the recommendation that the criteria not be used for market as proposed as they represent a potential threat to public health.

Population bioequivalence testing techniques proposed originally for the assessment of formulation equivalence are extended to ICH-E5 population pharmacokinetic testing to evaluate their use in that setting, and use of such approaches were not recommended in such trials.

In conclusion, as of the time of finalisation of this thesis, in part based upon findings published as part of the research in this thesis, the debate on individual and population bioequivalence has concluded in their removal from FDA regulatory guidance, but the use of alternative replicate designs for highly variable products is allowed for average bioequivalence testing. Further

developments in the area of ICH-E5 population equivalence testing will be a topic of future research.

References

[1] Aarons, L., Karlsson, M., Mentre, F., Rombout, F., Steimer, J.-L., van Peer, A., and Invited COST B15 Experts. (2001) The Role of Modelling and Simulation in Phase I Drug Development. *European Journal of Pharmaceutical Sciences*, **13**, 115–122.

[2] Akaike, H. (1973) Information Theory and An Extension of the Maximum Likelihood Pronciple. *Second International Symposium on Information Theory*, 267–281.

[3] Altman, D.G., Bland, J.M. (1995) Absence of evidence is not evidence of absence. *BMJ*, **311**, 485.

[4] Altman, J., Amit, O., Heise, M., Oldham, M., Patterson, S.D., Zariffa, N.M-D. (2000) ICH-E5: Guidance on Ethnic Factors in the Acceptability of Foreign Clinical Data, Points to Consider Document. *Biomedical and Data Sciences, GlaxoSmithKline Pharmaceuticals.*

[5] Anderson, S. (1993) Individual Bioequivalence: a problem of switchability [with discussion]. *Biopharm Rep*, **2**, 1–11.

[6] Anderson, S. (1995) Current issues in individual bioequivalence. *Drug Information Journal*, **29**, 961–964.

[7] Anderson, S., Hauck W.W. (1983) A new procedure for testing equivalence in comparative bioavailability and other clinical trials. *Comm Stat Theory Methods*, **12**, 2663–2692.

[8] Anderson, S., Hauck W.W. (1990) Consideration of Individual Bioequivalence. *J Pharmacokinet Biopharm*, **18**, 259–273.

[9] Anderson, S., Hauck W.W. (1996) The transitivity of bioequivalence testing: potential for drift. *Int J Clin Pharmacol Ther*, **34**, 369–374.

[10] Ansbacher, R. (1990) Interchangeability of low dose oral contraceptives. *Contraception*, **43**, 139–147.

[11] Balakrishnan, N., Ma, C.W. (1990) A comparative study of various tests for the equality of two population variances. *J Statist Comput Simul*, **35**, 41–89.

[12] Balthasar, J.P. (1999) Bioequivalence and Bioequivalency Testing. *American Journal of Pharmaceutical Education*, **63**, 194–198.

[13] Barbieri M.M., Liseo, B., Petrella, L. (2000) Bayes Factors for Fieller's Problem. *Biometrika*, **87**, 717–723.

[14] Barrett, J.S., Batra, V., Chow, A., Cook, J., Gould, A.L., Heller, A., Lo, MW., Patterson, S.D., Smith, B.P., Stritar, J.A., Vega, J.M., Zariffa, N. (2000) PhRMA Perspective on Population and Individual Bioequivalence and Update to the PhRMA Perspective on Population and Individual Bioequivalence. *J Clin Pharmacol*, **40**, 561–572.

[15] Basson, R.P., Ghosh, A., Cerimele, B.J., DeSante, K.A., Howey, D.C. (1998) Why rate of absorption inferences in single dose bioequivalence studies are often inappropriate. *Pharmaceutical Research*, **15**, 276–279.

[16] Bauer, P. (1991) Multiple Testing in Clinical Trials. *Statistics in Medicine*, **10**, 871–890.

[17] Bauer, P., Bauer, M.M. (1994) Testing equivalence simultaneously for location and dispersion of two normally distributed populations. *Biom J*, **6**, 643–660.

[18] Bauer, P., Kieser, M. (1996) A unifying approach for confidence intervals and testing of equivalence and difference. *Biometrika*, **83**, 934–937.

[19] Bekersky, I., Dressler, D., Colburn, W., Mekki, Q. (1999) Bioequivalence of 1 and 5 mg Tacrolimus Capsules using a Relicate Study Design. *Journal of Clinical Pharmacology*, **39**, 1032–1037.

[20] Bellavance, F., Tardif, S. (1995) A nonparametric approach to the analysis of three-treatment three-period cross-over data. *Biometrika*, **82**, 865–875.

[21] Benet, L.Z., Goyan, J.E. (1995) Bioequivalence and narrow therapeutic index drugs. *Pharmacotherapy*, **15**, 433–440.

[22] Benet, L.Z. (1999) Understanding Bioequivalence Testing. *Transplantation Proceedings*, **31, Suppl A**, 7S–9S.

[23] Berger, R.L. (1992) Multiparametric hypothesis testing and acceptance sampling. *Technometrics*, **24**, 295–300.

[24] Berger, R.L., Hsu, J.C. (1996) Bioequivalence trials, Intersection–Union Tests, and Equivalence Confidence Sets. *Statistical Science*, **11**, 283–319.

[25] Bhoj, D.S. (1979) Testing equality of variances of correlated variates with incomplete data on both responses. *Biometrika*, **66**, 681–683.

[26] Bickel, P.J., Doksum, K.A. (1977) *Mathematical Statistics.* Holden Day, SF.

[27] Biomarker Definition Working Group. (2001) Biomarkers and Surrogate Endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics*, **69**, 89–95.

[28] Blackwelder, W.C. (1982) Proving the null hypothesis in clinical trials. *Controlled Clin Trials*, **3**, 345–353.

[29] Blume, H.H., Midha, K.K., eds. (1993) Conference Report. In: Bio-International: Bioavailability, Bioequivalence, and Pharmacokinetics. *Medpharm Stuttgart*, 13–23.

[30] Boddy, A.W., Snikeris, F.C., Kringle, R.O., Wi, G.C-G., Oppermann, J.A., Midha, K.K. (1995) An approach for widening the bioequivalence acceptance limits in the case of highly variable drugs. *Pharmaceutical Research*, **12**, 1865–1868.

[31] Bois, F.Y., Tozer, T.N., Hauck, W.W., Chen, M.L., Patnaik, R., Williams, R. (1994) Bioequivalence: Performance of Several Measures of Extent of Absorption. *Pharmaceutical Research*, **11**, 715–722.

[32] Box, G.E., Cox, D.R. (1964) An analysis of transformations. *J R Stat Soc Series B*, **26**, 211–243.

[33] Breslow, N. with discussion (1990) Biostatistics and Bayes. *Statistical Sciene*, **5**, 269–298.

[34] Bristol, D.R. (1991a) Testing equality of treatment variances in a 2x2 cross-over study. *Journal of Biopharmaceutical Statistics*, **1**, 185–192.

[35] Bristol, D.R. (1991b) A confidence interval for the ratio of treatment variances in a 2x2 cross-over study. *Journal of Biopharmaceutical Statistics*, **1**, 237–245.

[36] Brown, B.W. (1980) The cross-over experiment for clinical trials. *Biometrics*, **36**, 69–79.

[37] Brown, E.B., Iyer, H.K., Wang, C.M. (1997) Tolerance intervals for assessing individual bioequivalence. *Statistics in Medicine*, **16**, 803–820.

[38] Brown, H.K., Kempton, R.A. (1994) The application of REML in clinical trials. *Statistics in Medicine*, **13**, 1601–1617.

[39] Brown, L.D., Hwang, J.T.G., Munk, A. (1997) An unbiased test for the bioequivalence problem. *Annals of Stat*, **25**, 2345–2367.

[40] Brown, M.B., Forsythe, A.B. (1974) Robust tests for the equality of variances. *Journal of the American Statistical Association*, **69**, 364–367.

[41] Buice, R.G., Subramanian, V.S., Duchin, K.L., Uko-Nne, S. (1996) Bioequivalence of a highly variable drug: An experience with Nadolol. *Pharmaceutical Research*, **13**, 1109–1115.

[42] Burdick, R.K., Graybill, F.A. (1992) *Confidence intervals on variance components.* Marcel Dekker, NY.

[43] Burdick, R.K., Sielken, R.L. (1978) Exact confidence intervals for linear combinations of variance components in nested classifications. *Journal of the American Statistical Association*, **73**, 632–635.

[44] Calvert, R.T. (1996) Bioequivalence and generic prescribing: a pharmacy view. *Journal of Pharmacy and Pharmacology*, **48**, 9–10.

[45] Canafax, D.M., Irish, W.D., Moran, H.B., Squiers, E., Levy, R., Pouletty, P., First, M.R., Christians, U. (1999) An Individual Bioequivalence Approach to Compare the Intra-subject Variability of Two Ciclosporin Formulations, SangCya and Neoral. *Pharmacology*, **59**, 78–88.

[46] Carter, B.L., Noyes, M.A., Demmler, R.W. (1993) Differences in serum concentrations of and responses to generic verapamil in the elderly. *Pharmacotherapy*, **13**, 359–368.

[47] Cartwright, A.C. (1991) International Harmonisation and Consensus DIA Meeting on Bioavailability and Bioequivalence Requirements and Standards. *Drug Information Journal*, **25**, 471–482.

[48] Casella, G., George, E.J. (1992) Explaining the Gibbs Sampler. *The American Statistician*, **46**, 167–174.

[49] Cerutti, R., Rivolta, G., Cavalieri, L., Di Giulio, C., Grossi, E., Vago, T., Baldi, G., Righini, V., Marzo, A. (1999) Bioequivalence of Levothyroxine Tablets administered to a target population in steady state. *Pharmacological Research*, **39**, 193–201.

[50] Chen, K-W., Li, G., Sun, Y. (1996) A Confidence Region Approach for Assessing Equivalence in Variability of Bioavailability. *Biom J*, **4**, 475–487.

[51] Chen, M-L. (1997) Individual Bioequivalence - a regulatory update. *Journal of Biopharmaceutical Statistics*, **7**, 5–11.

[52] Chen, M-L., Patnaik, R., Hauck, W.H., Schuirmann, D.J., Hyslop, T., Williams, R. (2000a) An individual bioequivalence criterion: regulatory considerations. *Statistics in Medicine*, **19**, 2821–2842.

[53] Chen, M-L., Lee, S-C., Ng, M-J., Schuirmann, D., Lesko, L.J., Williams, R.L. (2000b) Pharmacokinetic Analysis of Bioequivalence Trials: Implications for sex-related issues in clinical pharmacology and biopharmaceutics. *Clinical Pharmacology and Therapeutics*, **68**, 510–521.

[54] Chi, E.M. (1994) M-estimation in cross-over trials. *Biometrics*, **50**, 486–493.

[55] Chinchilli, V.M. (1996) The assessment of individual and population bioequivalence. *Journal of Biopharmaceutical Statistics*, **6**, 1–14.

[56] Chinchilli, V.M., Elswick, R.K. (1997) The multivariate assessment of bioequivalence. *Journal of Biopharmaceutical Statistics*, **7**, 113–123.

[57] Chinchilli, V.M., Esinhart, J.D. (1996) Design and analysis of intra-subject variability in cross-over experiments. *Statistics in Medicine*, **15**, 1619–1634.

[58] Chow, S.C., Liu, J. (2000) *Design and Analysis of Bioavailability and Bioequivalence Studies 2ed.* Marcel Dekker, NY.

[59] Chow, S-C., Shao, J., Wang, H. (2002) Individual Bioequivalence Testing under 2x3 designs. *Statist Med*, **21**, 629–48.

[60] Colaizzi, J.L., Lowenthal, D.T. (1986) Critical Therapeutic Ctegories: A Contraindication to Generic Substitiuation? *Clinical Therapeutics*, **8**, 370–379.

[61] Colburn, W.A., Keefe, D.L. (2000) Bioavailability and Bioequivalence: Average, Population, and/or Individual. *J Clin Pharmacol*, **40**, 559–600.

[62] Cornell, R.G. (1991) Nonparametric Tests of Dispersion for the Two Period Cross-over Design. *Commum Statist-Theory Meth*, **20**, 1099–1106.

[63] Cox, D.R. (1967) Fieller's theorem and a generalization. *Biometrika*, **54**, 567–572.

[64] Crowder, M.J., Kimber, A.C., Smith, R.L., Sweeting, T.J. (1991) *Statistical Analysis of Reliability Data*, Chapman and Hall, London.

[65] Crow, E.L., Shimizu, K. (1988) *Lognormal distributions*. Marcel Dekker, NY.

[66] D'Angelo, G., Potvin, D., Turgeon, J. (2001) Carry-over Effects in Bioequivalence Studies. *Journal of Biopharmaceutical Statistics*, **11**, 35–43.

[67] Davidian, M., Giltinan, D.M. (1995) *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall, London.

[68] Diletti, E., Hauschke, D., Steinijans, V.W. (1991) Sample size determination for bioequivalence assessment by means of confidence intervals. *Int J Clin Pharmacol Ther Toxicol*, **29**, 1–8.

[69] DiMasi, J. (2001) New Drug Development in the USA from 1963 to 1999; Risks in New Drug Development - Approval Success Rates for Investigational drugs. *Clinical Pharmacology and Therapeutics*, **69**, 286–307.

[70] Dragalin, V., Fedorov, V. (1999a) Kullback–Liebler Distance for evaluating bioequivalence. *SB BDS Technical Report 1999-03*.

[71] Dragalin, V., Fedorov, V. (1999b) The total least squares method in individual bioequivalence evaluation. *SB BDS Technical Report 1999-04*.

[72] Dragalin, V., Fedorov, V., Patterson, S., Jones, B. (2002) Kullback-Leibler divergence for evaluating bioequivalence. *Statistics in Medicine, In Press*.

[73] Dunnett, C.W., Gent, M. (1977) Significance testing to establish equivalence between treatments, with special reference to data in the form of 2 x 2 tables. *Biometrics*, **33**, 593–602.

[74] EFPIA-99: European Federation of Pharmaceutical Industries and Associations. (1999) Position Paper: Taiwan, Barriers to International Pharmaceuticals.

[75] Efron, B., Tibshirani, R.J. (1993) *An Introduction to the Bootstrap*. Chapman and Hall, NY.

[76] Ekbohm, E. (1981) A test for the equality of variances in the paired case with incomplete data. *Biom J*, **3**, 261–265.

[77] Ekbohm, G., Melander, H. (1989) The subject-by-formulation interaction as a criterion of interchangeability of drugs. *Biometrics*, **45**, 1249–1254.

[78] El-Tahtawy, A.A., Tozer, T.N., Harrison, F., Lesko, L., Williams, R. (1998) Evaluation of bioequivalence of highly variable drug using clinical trial simulations. II: Comparison of Single and Multiple-Dose Trials using AUC and Cmax. *Pharmaceutical Research*, **15**, 98–104.

[79] European Agency for the Evaluation of Medicinal Products, Committee for Proprietary Medicinal Products. (2001) Draft Guidance on the Investigation of Bioavailability and Bioequivalence.

[80] Endrenyi, L., Fritsch, S., Yan, W. (1991) Cmax/AUC is a clearer measure than Cmax for absorption rates in investigations of bioequivalence. *International Journal of Clinical Pharmacology, Therapy, and Toxicology*, **29**, 394–399.

[81] Endrenyi, L., Schulz, M. (1993) Individual variation and the acceptance of average bioequivalence. *Drug Information Journal*, **27**, 195–201.

[82] Endrenyi, L. (1994) A method for the evaluation of individual bioequivalence. *Int J Clin Pharmacol Ther*, **32**, 497–508.

[83] Endrenyi, L., Czimadia, F., Tothfalusi, L, Chen, M.L. (1998a) Metrics comparing simulated early concentration profiles for the determination of bioequivalence. *Pharmaceutical Research*, **15**, 1292–1299.

[84] Endrenyi, L., Amidon, G.L., Midha, K.K., Skelly, J.P. (1998b) Individual bioequivalence: attractive in principle, difficult in practice. *Pharmaceutical Research*, **15**, 1321–1325.

[85] Endrenyi, L., Hao, Y. (1998c) Assymetry of the mean-variability tradeoff raises questions about the model in investigations of individual bioequivalence. *Int J Clin Pharmacol Ther Toxicol*, **36**, 1–8.

[86] Endrenyi, L., Midha, K.K. (1998d) Individual bioequivalence–has its time come? *European Journal of Pharmaceutical Sciences*, **6**, 271–277.

[87] Endrenyi, L., Taback, N., Tothfalusi, L. (2000) Properties of the Estimated Variance Component for Subject-by-formulation interaction in Studies of Individual Bioequivalence. *Statistics in Medicine*, **19**, 2867–2878.

[88] Endrenyi, L., Tothfalusi, L. (1999) Subject-by-formulation interaction in determination of individual bioequivalence: bias and prevalence. *Pharmaceutical Research*, **16**, 186–190.

[89] Ereshefsky, L., Meyer, M. (2001) Comparison of the Bioequivalence of Generic versus Branded Clozapine. *J Clin Psychiatry*, **62(Suppl 5)**, 3–27.

[90] Esinhart, J.D., Chinchilli, V.M. (1994a) Extensions to the use of tolerance intervals for the assessment of individual bioequivalence. *Journal of Biopharmaceutical Statistics*, **4**, 39–52.

[91] Esinhart, J.D., Chinchilli, V.M. (1994b) Sample size considerations for assessing individual bioequivalence based on the method of tolerance intervals. *Int J Clin Pharmacol Ther*, **32**, 26–32.

[92] EFPIA-99: European Federation of Pharmaceutical Industries and Associations. (1999) Position Paper: Taiwan, Barriers to International Pharmaceuticals.

[93] FDA Guidance. (1992) Statistical procedures for bioequivalence studies using a standard two treatment cross-over design.

[94] FDA Guidance. (1995) Guidance for Industry: Immediate Release Solid Oral Dosage Forms Scale-Up and Post-Approval Changes: Chemistry, Manufacturing and Controls, In Vitro Dissolution Testing, and In Vivo Bioequivalence Documentation.

[95] FDA Draft Guidance. (1997) In vivo bioequivalence studies based on population and individual bioequivalence approaches.

[96] FDA Draft Guidance. (1999a) BA and BE Studies for Orally Administered Drug Products: General Considerations.

[97] FDA Draft Guidance. (1999b) Average, Population, and Individual Approaches to Establishing Bioequivalence.

[98] FDA Guidance. (1999c) Population Pharmacokinetics.

[99] FDA Guidance. (2000a) Waiver of In Vivo Bioavailability and Bioequivalence Studies for Immediate-Release Solid Oral Dosage Forms Based on a Biopharmaceutics Classification System.

[100] FDA Guidance. (2000b) Bioavailability and Bioequivalence Studies for Orally Administered Drug Products: General Considerations.

[101] FDA Guidance. (2001) Statistical Approaches to Establishing Bioequivalence.

[102] FDA Draft Guidance. (2002) Bioavailability and Bioequivalence Studies for Orally Administered Drug Products: General Considerations.

[103] Federer, W.T., Raghavarao, D. (1975) On augmented designs. *Biometrics*, **31**, 29–35.

[104] Fieller, E. (1954) Some problems in interval estimation. *J R Stat Soc B*, **16**, 175–185.

[105] Fisher, L.D., van Belle, G. (1993) *Biostatistics, A Methodology for the Health Sciences.* John Wiley and Sons, New York.

[106] Fleiss, J.L. (1971) On the distribution of a linear combination of independent chi squares. *Journal of the American Statistical Association*, **66**, 142–144.

[107] Fleiss, J.L. (1986) Letter to the Editor: On Multiperiod Cross-over Studies. *Biometrics*, 449–450.

[108] Fleiss, J.L. (1989) A critique of recent research on the two treatment cross-over design. *Controlled Clinical Trials*, **10**, 237–243.

[109] Fluehler, H., Hirtz, J., Moser, H.A. (1981) An aid to decision making in bioequivalence assessment. *J Pharmacokinet Biopharm*, **9**, 235–243.

[110] Freeman, P.R. (1989) The performance of the two-stage analysis of two-treatment, two-period cross-over trials. *Statistics in Medicine*, **8**, 1421–32.

[111] Gaffney, M. (1992) Variance components in comparative bioavailability studies. *Journal of Pharmaceutical Sciences*, **81**, 315–317.

[112] Gelfland, A.E., Hills, S.E., Racine-Poon, A., Smith, A.F-M. (1990) Illustration of Bayesian Inference in Normal Data Models using Gibbs Sampling. *Journal of the American Statistical Association*, **85**, 972–985.

[113] Gelfand, A.E., Smith, A.F-M. (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.

[114] Ghosh, K., Hesney, M., Sarkar, S., Thiyagarajan, B. (1996) Comprehensive Criteria for Population and Individual Bioequivalence. *Proceedings of the American Statistical Association, Biopharmaceutics Chapter*, 77-81.

[115] Giesbrecht, F., Burns, J. (1985) Two-stage analysis based on a mixed model: Large sample asymptotic theory and small-sample simulation results. *Biometrics*, **41**, 477–486.

[116] Godbillon, J., Cardot, J.B., LeCaillon, G., Sioufi, A. (1996) Bioequivalence assessment: a pharmaceutical industry perspective. *European Journal of Drug Metabolism and Pharmacokinetics*, **21**, 153–158.

[117] Goodman, S. (1994) The use of predictive confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med*, **121**, 200–206.

[118] Gould, A.L. (1995) Group sequential extensions of a standard bioequivalence testing procedure. *J Pharmacokin Biopharm*, **23**, 57–86.

[119] Gould, A.L. (1997) Discussion of Individual Bioequivalence by ML Chen. *Journal of Biopharmaceutical Statistics*, **7**, 23–29.

[120] Gould, A.L. (2000a) A practical approach for evaluating population and individual bioequivalence. *Statistics in Medicine*, **19**, 2721–2740.

[121] Gould, A.L., Li, S.X., Bradstreet, T.E. (2000b) Evaluating Individual Bioequivalence. *Presented at the International Biometrics Conference, Berkley, CA*, **July**.

[122] Grahnan, A., Hammarlund, M., Lundqvist, T. (1984) Implications of intraindividual variability in bioavailability studies of furosemide. *Eur J Clin Pharmacol*, **27**, 595–602.

[123] Graybill, F.A., Wang, CH. (1980) Confidence intervals on nonnegative linear combinations of variances. *Journal of the American Statistical Association*, **75**, 869–873.

[124] Grieve, A.P. (1985) A Bayesian analysis of the two-period cross-over design for clinical trials. *Biometrics*, **21**, 467–480.

[125] Grizzle, J.E. (1965) The two-period changeover design and its use in clinical trials. *Biometrics*, **21**, 467–480.

[126] Guilbaud, O. (1993) Exact inference about the within-subject variability in 2 x 2 cross-over studies. *Journal of the American Statistical Association*, **88**, 939–946.

[127] Guilbaud, O. (1999) Exact comparisons of means and variances in 2x2 cross-over trials. *Drug Information Journal*, **33**, 455–469.

[128] Hare, D., Foster, T. (1990) The Orange Book: the FDA's advice on therapeutic equivalence. *American Pharmacy*, **NS30**, 35–37.

[129] Harville, D.A. (1976) Confidence intervals and sets for linear combinations of fixed and random effects. *Biometrics*, **32**, 403–407.

[130] Harville, D.A. (1977) Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72**, 320–340.

[131] Harville, D.A. (1990) BLUP (Best Linear Unbiased Prediction) and Beyond. *AdStMetGen*, 239–276.

[132] Harville, D.A., Carriquiry, A.L. (1992) Classical and Bayesian Prediction as Applied to an Unbalanced Mixed Linear Model. *Biometrics*, **48**, 987–1003.

[133] Harville, D.A., Jeske, D.R. (1992) Mean Squared Error of Estimation or Prediction under a General Linear Model. *Journal of the American Statistical Association*, **87**, 724–731.

[134] Hauck, W.W., Anderson S. (1984) A new statistical procedure for testing equivalence in two-group comparative bioavailability trials. *J Pharmacokin Biopharm*, **12**, 83–91.

[135] Hauck, W.W., Anderson S. (1986) A proposal for interpreting and reporting negative studies. *Statistics in Medicine*, **5**, 203–209.

[136] Hauck, W.W., Anderson S. (1992) Types of bioequivalence and related statistical considerations. *Int J Clin Pharmacol Ther Toxicol*, **30**, 181–187.

[137] Hauck, W.W., Anderson S. (1994) Measuring switchability and presciptability: when is average bioequivalence sufficient. *J Pharmacokinet Biopharm*, **22**, 551–564.

[138] Hauck, W.W, Bois, F.Y., Hyslop, T., Gee, L., Anderson, S. (1997) A parametric approach to population bioequivalence. *Statistics in Medicine*, **16**, 441–454.

[139] Hauck, W.W, Chen, ML., Hyslop, T., Patnaik, P., Shuirmann, D., Williams, R., for the FDA Individual Bioequivalence Working Group. (1996) Mean differences versus variability reduction: tradeoff in aggregate measures for individual bioequivalence. *Int J Clin Pharmacol Ther Toxicol*, **34**, 535–541.

[140] Hauck, W.W., Hauschke, D., Diletti, E., Bois, F.Y., Steinijans, V.W., Anderson, S. (1997) Choice of student's t or wilcoxon based confidence intervals for assessment of bioequivalence. *Journal of Biopharmaceutical Statistics*, **7**, 179–189.

[141] Hauck, W.W., Hyslop, T., Anderson, S., Bois, F.Y., Tozer, T.N. (1995) Statistical and regulatory considerations for multiple measures in bioequivalence testing. *Clin Res Reg Affairs*, **12**, 249–265.

[142] Hauck, W.W., Preston, P.E., Bois, F.Y. (1997) A group sequential approach to cross-over trials for average bioequivalence. *Journal of Biopharmaceutical Statistics*, **7**, 87–96.

[143] Hauck, W.W., Hyslop, T., Chen, M.L., Patnaik, R., Williams, R.L., and the FDA Population/Individual Bioequivalence Working Group. (2000) Subject-by-Formulation Interaction in Bioequivalence: Conceptual and Statistical Issues. *Pharmaceutical Research*, **17**, 375–380.

[144] Hauschke, D., Steinijans, V.W., Diletti, E. (1990) A distribution free procedure for the statistical analysis of bioequivalence studies. *Int J Clin Pharmacol Ther Toxicol*, **28**, 72–78.

[145] Hauschke, D., Steinijans, V.W. (2000) The US draft guidance regarding population and individual bioequivalence approaches: comments by a research-based pharmaceutical company. *Statistics in Medicine*, **19**, 2769–2774.

[146] Hauschke, D. (2002) Letter to the Editor and Correspondence on 'A Note on Sample Size Calculation in Bioequivalence Trials' by Chow and Wang (2001;28:155-169). *Journal of Pharmacokinetics and Pharmacodynamics*, **29**, 89–102.

[147] Haynes, J.D. (1981) Statistical simulation study of new proposed uniformity requirement for bioequivalency studies. *Journal of Pharmaceutical Sciences*, **70**, 673–675.

[148] Hellriegel, E.T., Bjornsson, T.D., Hauck, W.W. (1996) Interpatient variability in bioavailability is related to the extent of absorption: implications for bioavailability and bioequivalence studies. *Clinincal Pharmacology and Therapeutics*, **60**, 601–607.

[149] Herchuelz, A. (1996) Bioequivalence assessment and the conduct of bioequivalence trials: a European point of view. *European Journal of Drug Metabolism and Pharmacokinetics*, **21**, 149–152.

[150] Hills, M., Armitage, P. (1979) The two period cross-over trial. *Br J Clin Pharmacol*, **8**, 7–20.

[151] Hinkelmann, K., Kempthorne, O. (1994) *Design and Analysis of Experiments, Volume I: Introduction to Experimental Design*, John Wiley and Sons, NY.

[152] Hoenig, J., Heisey, D. (2001) The Abuse of Power: The pervasive fallacy of power calculations for data analysis. *The American Statistician*, **55**, 19–24.

[153] Holder, D. (1993) PhD Dissertation, Moment based criteria for bioequivalence. *Temple University.*

[154] Holder, D., Hsuan, F. (1993a) Moment based criteria for determining bioequivalence. *Biometrika*, **80**, 835–846.

[155] Holder, D., Hsuan, F. (1993b) A Moment based criteria for determining individual bioequivalence. *Drug Information Journal*, **29**, 965–979.

[156] Howe, W.G. (1974) Approximate confidence limits on the mean of X+Y where X and Y are two tabled independent random variables. *Journal of the American Statistical Association*, **69**, 789–794.

[157] Hsu, J.C., Hwang, J-T.G., Liu, H-K., Ruberg, S.J. (1994) Confidence limits associated with tests for bioequivalence. *Biometrika*, **81**, 103–114.

[158] Huitson, A. (1955) A method for assigning confidence limits to linear combinations of variances. *Biometrika*, **42**, 471–479.

[159] Hunt, M.I. (2000) Prescription Drugs and Intellectual Property Protection. *NIHCM, National Istitute for Health Care Management: Foundation Issue Brief.*

[160] Huque, M.F., Dubey, S., Fredd, S. (1989) Establishing Therapeutic Equivalence with Clinical Endpoints. *Proceedings of the American Statistical Association, Biopharmaceutics Chapter*, 46–52.

[161] Huynh, H., Feldt, L.S. (1970) Conditions under which mean square ratios in repeated measures designs have exact F-distributions. *Journal of the American Statistical Association*, **65**, 1582–1589.

[162] Hwang, J.S. (1996) Numerical Solutions for a Sequential Approach to Bioequivalence. *Statistica Sinica*, **6**, 663–673.

[163] Hwang, J.T.G., Wang, W. (1997) The validity of the test of individual equivalence ratios. *Biometrika*, **84**, 893–900.

[164] Hwang, S., Huber, P.B., Hesney, M., Kwan, K.C. (1978) Bioequivalence and interchangeabiliy. *Journal of Pharmaceutical Sciences*, **67**, IV.

[165] Hyslop, T., Hsuan, F., Holder, D.J. (1999) A small sample confidence interval approach to assess individual bioequivalence. *Presented at 1999 Biostatistics and Clinical Data Management Workshop.*

[166] Hyslop, T., Hsuan, F., Holder, D.J. (2000) A small sample confidence interval approach to assess individual bioequivalence. *Statistics in Medicine*, **19**, 2885–2897.

[167] Hyslop, T., Inglewicz, B. (2001) Alternative Cross-over Designs for Individual Bioequivalence. *Proceedings of the Annual Meeting of the American Statisticial Association*, August 5-9.

[168] International Conference on Harmonization. (1994) Dose Response Information to Support Drug Registration.

[169] International Conference on Harmonization. (1998) Guidance on Ethnic Factors in the Acceptability of Foreign Clinical Data.

[170] Jennison, C., Turnbull, B.W. (2000) Group Sequential Methods with Applications to Clinical Trials. Chapman and Hall.

[171] Jennrich, R.I., Schluchter, M.D. (1986) Unbalanced Repeated-Measures Models with Structured Covariance Matrices. *Biometrics*, **42**, 805–820.

[172] Johnson, N.L., Kotz, S., Balakrishnan, N. (1994) *Continuous Univariate Distributions Vol 1 and 2.* John Wiley and Sons, NY.

[173] Jones, B., Kenward M.G. (1989) *Design and Analysis of Cross-over Trials.* Chapman and Hall, London.

[174] Jones, B., Jarvis, P., Lewis, J.A., Ebbutt, A.F. (1996) Trials to assess equivalence: the importance of rigourous methods. *BMJ*, **313**, 36–39.

[175] Jones, B., Deppe, C. (2000) Recent Developments in the Design of Cross-over trials: A brief review and bibliography. *Proceedings of the Conference on Recent Developments in the Design of Experiments and Related Topics*, Nova Science.

[176] Jones, B., Patterson, S.D. (2002) Tutorial: Design and Analysis of Bioequivalence Trials. *Drug Information Agency 2002 Annual Meeting.*

[177] Julious, S., Debarnot, C. A-M. (2000) Why are pharmacokinetic data summarized by arithmetic means? *Journal of Biopharmaceutical Statistics*, **10**, 55–71.

[178] Kanfer, F-H.J., Geertsema, J.C., Steyn, H.S. (1988) A Two-Stage Procedure for Testing Bioequivalence. *Proceedings of the American Statistical Association*, **Biopharmaceutical Chapter**, 148–152.

[179] Kawai, N., Andoh, M., Uwoi, T., Goto, M. (2000) Statistical Approaches to Accepting Foreign Clinical Data. *Drug Information Journal*, **34**, 1265–1272.

[180] Kenward, M., Roger, J. (1997) Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics*, **33**, 983–997.

[181] Kershner, R.P., Federer, W.T. (1981) Two treatment cross-over designs for estimating a variety of effects. *Journal of the American Statistical Association*, **76**, 612–619.

[182] Kepner, J.L., Randles, R.H. (1982) Detecting unequal marginal scales in a bi-variate population. *Journal of the American Statistical Association*, **77**, 475–482.

[183] Kimanani, E.K., Potvin, D. (1997) A parametric confidence interval for a moment based scaled criterion for individual bioequivalence. *J Pharmacokinet Biopharm*, **25**, 595–614.

[184] Kimanani, E.K., Lavigne, J., Potvin, D. (2000a) Numerical methods for the evaluation of individual bioequivalence. *Statistics in Medicine*, **19**, 2775–2795.

[185] Kimanani, E., Stypinski, D., Curtis, G., Stiles, M., Heessels, P., Logan, S., Melson, K., St Germain, E., Boswell, G. (2000b) A Contract Research Organization's Response to the New FDA Guidances for Bioequivalence/Bioavailability Studies for Orally Administered Drug Products. *Journal of Clinical Pharmacology*, **40**, 1102–1108.

[186] Kirkwood, T-B.L., Westlake, T.J. (1981) Letter to the Editor and Response. *Biometrics*, **37**, 589–593.

[187] Kullback, S. (1968) *Information Theory and Statistics.* Dover Publications, NY.

[188] Lacey, L.F., Keene, O.N., Bye, A. (1995) Glaxo's expreience of different absorption rate metrics of immediate release and extended release dosage forms. *Drug Information Journal*, **29**, 821–840.

[189] Lacey, L.F., Keene, O.N., Pritchard, J.F., Bye, A. (1997) Common Noncompartmental Pharmacokinetic Variables: are they normally or log-normally distributed? *Journal of Biopharmaceutical Statistics*, **7**, 171–178.

[190] Laird, N.M., Ware, J.H. (1982) Random effects models for longitudinal data. *Biometrics*, **38**, 963–974.

[191] Lesko, L., Rowland, M., Peck, C., Blaschke, T. (2000) Optimizing the Science of Drug Development: opportunities for better candidate selection and accelerated evaluation in humans. *European Journal of Pharmaceutical Sciences*, **10**, iv–xiv.

[192] Levy, G. (1995) The clay feet of bioequivalence testing. *Journal of Pharmacy and Pharmacology*, **47**, 975–977.

[193] Lin, L. (1989) A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, **45**, 255–268.

[194] Lin, L. (1992) Assay validation using the concordance correlation coefficient. *Biometrics*, **48**, 599–604.

[195] Lin, L. (2000) Total Deviation Index for Measuring Individual Agreement with Applications in Laboratory Performance and Bioequivalence. *Statistics in Medicine*, **19**, 255–270.

[196] Lindley, D.V. (1971) *Making Decisions 2ed.* John Wiley and Sons, NY.

[197] Lindley, D.V, Singpurwalla, N.D. (1991) On the evidence needed to reach agreed action between adversaries, with application to acceptance sampling. *Journal of the American Statistical Association*, **86**, 933–937.

[198] Lindley, D.V. (1997) The choice of sample size (with discussion). *The Statistician*, **46**, 129–166.

[199] Lindley, D.V. (1998) Decision analysis and bioequivalence trials. *Statistical Science*, **13**, 136–141.

[200] Lindsey, J.K. (1996) *Parametric Statistical Inference.* Clarendon Press, Oxford.

[201] Lindsey, J.K., Jones, B. (1997) Treatment-Patient Interactions for Diagnostics of Cross-over Trials. *Statistics in Medicine*, **16**, 1955–1964.

[202] Lindsey, J.K., Wang, J., Byrom, W.D., Jones, B. (1999) Modeling the Covariance Structure in Pharmacokinetic Cross-over Trials. *Journal of Bipharmaceutical Statistics*, **9**, 439–450.

[203] Lindstrom, M.J., Bates, D.M. (1988) Newton-Raphson and EM algorithms for Linear Mixed-Effects Models and Repeated-Measures Data. *Journal of the American Statisticial Association*, **83**, 1014–1022.

[204] Liu, J-P. (2000) Presentation: Methodological Issues in Evaluation of Bridging Evidence for Pharmaceutical Products. *First Workshop on APEC Bridging, Taipei, Taiwan, ROC.*

[205] Locke, C.S. (1984) An exact confidence interval for untransformed data for the ratio of two formulation means. *J Pharmacokinet Biopharm*, **12**, 649–655.

[206] Longford, N.T. (1999) Selection bias and treatment heterogeneity in clinical trials. *Statistics in Medicine*, **18**, 1467–1474.

[207] Longford, N.T. (2000) An alternative definition of individual bioequivalence. *Statistica Neerlandica*, **14**, 14–36.

[208] Lu, T.C., Graybill, F.A., Burdick, R.K. (1988) Confidence intervals on a difference of expected mean squares. *J Stat Plan Inf*, **18**, 35–43.

[209] Lund, R.E. (1975) Tables for an approximate test for outliers in linear models. *Technometrics*, **17**, 473–476.

[210] Machado, S., Miller, R., Hu, C. (1999) A regulatory perspective on pharmacokinetic and pharmacodynamic modelling. *Statistical Methods in Medical Research*, **8**, 217–45.

[211] Mallet, A. (1986) A maximum likelihood estimation method for random coefficient regression models. *Biometrika*, **73**, 645–56.

[212] Mancinelli, L., Frassetto, L., Floren, L., Dressler, D., Carrier, S., Bekersky, I., Benet, L., Christians, U. (2001) The Pharmacokinetics and Metabolic Disposition of Tacrolimus: A comparison across ethnic groups. *Clinical Pharmacology and Therapeutics*, **69**, 24–31.

[213] Mandallaz, D., Mau, J. (1981) Comparison of different methods for decison making in bioequivalence assessment. *Biometrics*, **37**, 213–222.

[214] Marcus, R., Peritz, E., Gabriel, K.R. (1976) On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, **63**, 655–60.

[215] Marston, S.A., Polli, J.E. (1997) Evaluation of direct curve comparison metrics applied to pharmacokinetic profies and relative bioavailability and bioequivalence. *Pharmaceutical Research*, **14**, 1363–1369.

[216] Marzo, A., Monti, N.C., Tettamanti, R.A., Crivelli, F., Bo, L.D., Mazzucchelli, P., Meoli, A., Pezzuto, D., Corsico, A. (2000) Bioequivalence of Inhaled Formoterol Fumurate Assessed from Pharmacodynamic, Safety, and Urinary Pharmacokinetic Data. *Arzneim-Forsch./Drug Res.*, **50**, 559–563.

[217] Mauger, D.T., Chinchilli, V.M. (2000) An alternative index for assessing profile similarity in bioequivalence trials. *Statistics in Medicine*, **19**, 2855–2866.

[218] McCulloch, C.E. (1987) Tests for equality of variances with paired data. *Commun Stat Theory Methods*, **16**, 1377–1391.

[219] Metzler, C.M. (1974) Bioavailability: a problem in equivalence. *Biometrics*, **30**, 309–317.

[220] Metzler, C.M., Huang, D.C. (1983) Statistical methods for bioavailability and bioequivalence. *Clin Res Pract Drug Reg Affairs*, **1**, 109–132.

[221] Meyer, M.C. (1995) Current scientific issues regarding bioavailability/bioequivalence studies: an academic view. *Drug Information Journal*, **29**, 805–812.

[222] Meyer, M.C., Straughn, A.B., Jarvi, E.J., Patrick, K.S., Pelsor, F.R., Williams, R.L., Patnaik, R., Chen, M.L., Shah, V.P. (2000) Bioequivalence of Methylphenidate Immediate-Release Tablets uing a Replicated Study Design to Characterize Intrasubect Variability. *Pharmaceutical Research*, **17**, 381–384.

[223] Meyer, M.C., Straughn, A.B., Mhatre, R.M., Shah, V.P., Chen, M-L., Williams, R.L., Lesko, L.J. (2001) Variability in the Bioavailability of Phenytoin Capsules in Males and Females. *Pharmaceutical Research*, **18**, 394–397.

[224] Midha, K.K., Ormsby, E.D., Hubbard, J.W., McKay, G., Hawes, E.M., Gavalas, L., McGilveray, I.J. (1993) Logarithmic Transformation in Bioequivalence: Application with Two Formulations of Perphenazine. *Journal of Pharmaceutical Sciences*, **82**, 138–144.

[225] Midha, K.K., Rawson, M.J., Hubbard, J.W. (1997a) Individual and average bioequivalence of high variability drugs and drug products. *Journal of Pharmaceutical Sciences*, **86**, 1193–1197.

[226] Midha, K.K., Rawson, M.J., Hubbard, J.W. (1997b) Bioequivalence: switchability and scaling. *European Journal of Pharmaceutical Sciences*, **6**, 87–91.

[227] Miller, M. (2001) Gender-Based Differences in the Toxicity of Pharmaceuticals - The Food and Drug Administration's Perspective. *International Journal of Toxicology*, **20**, 149–152.

[228] Milliken, G.A., Johnson, D.E. (1992) *Analysis of Messy Data, Vol 1: Designed Experiments*, Chapman and Hall, NY.

[229] Morgan, W.A. (1939) A test for the significance of the difference between two variances in a sample from a bi-variate normal population. *Biometrika*, **31**, 13–19.

[230] Muirhead, R.J. (1982) *Aspects of Multivariate Statistical Theory*. John Wiley and Sons, NY.

[231] Munk, A. (1993) An improvement on commonly used tests in bioequivalence assessment. *Biometrics*, **49**, 1225–1230.

[232] Munk, A., Pfluger, R. (1999) 1-α equivalence confidence rules for convex alternatives are $\alpha/2$ tests with applications to the multivariate assessment of bioequivalence. *Journal of the American Statistical Association*, **94**, 1311–1319.

[233] Nagata, R., Fukase, H., Rafizadeh-Kabe, J. (2000) East-West development: understanding the usability and acceptance of foreign data in Japan. *International Journal of Clinical Pharmacology and Therapeutics*, **38**, 87–92.

[234] Naito, C. (1994a) Ethnic Factors in the Acceptability of Foreign Data, Rapporteur's Introduction. *Proceedings of the Second International Conference on Harmonisation, 1993*, D'Arcy, PF., Harron, D. eds, Greystone Books, Ltd.

[235] Naito, C. (1994b) Ethnic Effects on Pharmacokinetic Parameters. *The Relevance of Ethnic Factors in the Clinical Evaluation of Medicines*, Walker, S., Lumley, C., McAnslane, N. eds, Kluwer Academic Publishers.

[236] Naito, C. (1998a) Relevance of Ethnic Factors in Global Development, Historic Overview. *Proceedings of the Fourth International Conference on Harmonisation, 1997*, D'Arcy, PF., Harron, D. eds, Greystone Books, Ltd.

[237] Naito, C. (1998b) Ethnic Factors in the Acceptability of Foreign Clinical Data. *Drug Information Journal*, **32**, 1283S–1292S.

[238] O'Brien, P. (1988) Comparing two samples: extensions of the t, rank-sum, and log-rank tests. *Journal of the American Statistical Association*, **83**, 52–61.

[239] Ormsby, E.D. (1999) Invited Presentation: Individual Bioequivalence: A Canadian Perspective. *AAPS International workshop on Individual Bioequivalence: Realities and Implementation.*

[240] O'Quigley, J., Baudoin, C. (1988) General approaches to the problem of bioequivalence. *The Statistician*, **37**, 51–58.

[241] Owen, D.B. (1965) A special case of a bi-variate non-central t-distribution. *Biometrika*, **52**, 437–446.

[242] Pabst, G., Jaeger, H. (1990) Review of methods and criteria for the evaluation of bioequivalence studies. *Eur J Clin Pharmacol*, **38**, 5–10.

[243] Patnaik, P.B. (1949) The non-central χ^2 and F-distributions and their applications. *Biometrika*, **36**, 202–232.

[244] Patnaik, R., Lesko, L.J., Chan, K., Williams, R.L. (1996) Bioequivalence assessment of generic drugs: an American point of view. *European Journal of Drug Metabolism and Pharmacokinetics*, **21**, 159–164.

[245] Patnaik, R.N., Lesko, L.J., Chen, ML., Williams, R.J., and the FDA Individual Bioequivalence Working Group. (1997) Individual Bioequivalence–new concepts in the statistical assessment of bioequivalence metrics. *Clin Pharmacokinet*, **33**, 1–6.

[246] Patterson, H.D. (1950) The construction of balanced designs for experiments involving sequences of treatments. *Biometrika*, **39**, 32.

[247] Patterson, H.D., Thompson, R. (1971) Recovery of Inter-block information when block sizes are unequal. *Biometrika*, **58**, 545–554.

[248] Patterson, S.D., Zariffa, N.M–D., Boyle, D., Hyneck, M. (1998) Poster presentation: Case studies, Practical Issues, and Observations on Population and Individual Bioequivalence. *Annual Midwest Biopharmaceutics Workshop.*

[249] Patterson, S.D., Zariffa, N.M–D. (1999) Invited presentation: An innovator industry perspective on application and experiences with Population and Individual bioequivalence. *AAPS International workshop on Individual Bioequivalence: Realities and Implementation.*

[250] Patterson, S.D., Zariffa, N.M–D. (2000a) Poster presentation: Case studies in the measurement of agreement between regimens in cross-over studies. *Annual Meeting of the American Society of Clinical Pharmacology and Therapeutics.*

[251] Patterson, S.D., Zariffa, N.M–D., Montague, T., Howland, K. (2000b) Approaches to meeting a nearly impossible regulatory hurdle: Demonstrating average bioequivalence for a highly variable drug product. *Annual Meeting of the American Society of Clinical Pharmacology and Therapeutics.*

[252] Patterson, S. (2000c) A Statistical Review of Bioequivalence: Concepts, Controversies, and Historical Perspectives. *Presented at University College, London Journal Club.*

[253] Patterson, S.D., Leonov, S., Mathur, A., Sparrow, P., Zhu, J. (2000d) BDS Technical Policy: Design and Analysis of Dose Response Studies. *SmithKline Beecham Pharmaceuticals, Biostatistics and Data Sciences Technical Policy Document.*

[254] Patterson, S. (2001a) A Statistical Review of Bioequivalence: Concepts, Controversies, and Historical Perspectives. *Presented at Barlett Internation Conference on Bioavailability and Bioequivalence.*

[255] Patterson, S. (2001b) A Statistical Review of Bioequivalence: Concepts, Controversies, and Historical Perspectives. *Presented at American Statistical Association Philadelphia Chapter.*

[256] Patterson, S. (2001c) A Review of the Development of Biostatistical Design and Analysis Techniques for Assessing In Vivo Bioequivalence, Part 1. *Indian Journal of Pharmaceutical Sciences*, **63**, 81–100.

[257] Patterson, S. (2001d) A Review of the Development of Biostatistical Design and Analysis Techniques for Assessing In Vivo Bioequivalence, Part 2. *Indian Journal of Pharmaceutical Sciences*, **63**, 169–186.

[258] Patterson, S., Dragalin, V., Fedorov, V., Jones, B. (2001e) Kullback–Liebler Distance for evaluating bioequivalence II. Simulation and Retrospective Assessment of Performance of the Proposed FDA metric and the Kullback-Leibler Divergence in Individual Bioequivalence Assessment. *GSK BDS Technical Report 2001-02.*

[259] Patterson, S., Jones, B., Zariffa, N. (2001f) Population Bioequivalence and the Evaluation of Ethnic Factors in the Acceptance of Equivalence Data. *Presented at the Societe de Statistique Francaise.*

[260] Patterson, S., Jones, B. (2001g) Bioequivalence and the Pharmaceutical Industry. *Presented at the Annual Meeting of the Statisticians in the Pharmaceutical Industry (PSI).*

[261] Patterson, S., Zariffa, N., Howland, K., Montague, T. (2001h) Non-traditional study designs to demonstrate average bioequivalence for highly variable drug products. *Eur J Clin Pharmacol*, **57**, 663–670.

[262] Patterson, S., Amit, O., Jones, B., Zariffa, N. (2001i) Presentation: Bridging Strategies and ICH-E5: How to realise the potential of ICH-E5 in Drug Development. *Drug Information Agency Annual Meeting.*

[263] Patterson, S., Jones, B. (2002a) Bioequivalence and the Pharmaceutical Industry. *Pharmaceutical Statistics*, IN PRESS.

[264] Patterson, S., Jones, B. (2002b) Replicate Designs and Average, Individual, and Population Bioequivalence: I. Estimation, Inference, and Retrospective Assessment of Performance of Novel Procedures and the Proposed FDA Methods for Bioequivalence Assessment. *GSK BDS Technical Report 2002-01.*

[265] Patterson, S., Jones, B. (2002c) Replicate Designs and Average, Individual, and Population Bioequivalence: II. Simulation Study of Performance of Novel Procedures and the Proposed FDA Methods for Bioequivalence Assessment. *GSK BDS Technical Report 2002-05.*

[266] Patterson, S., Altman, J., Asbel, N., Barrows, C., Dragalin, V., Giguere, G., Graham, A., Hale, M., Keene, O., Lettis, S., Quinlan, J., Segreti, T. (2002d) Statistical Points to Consider in Proof-of-Concept: Design, Analysis, Interpretation, and Presentation. *GlaxoSmithKline Pharmaceuticals Biomedical and Data Sciences Report.*

[267] Patterson, S., Jones, B. (2002e) Statistical Aspects of Bioequivalence in the Pharmaceutical Industry. *Presented at the American Statistical Association Joint Statistical Meetings.*

[268] Patterson, S., Jones, B. (2002f) REML and Method-of-Moments Modelling as Applied to Replicate Cross-over Designs - Implications for Bioequivalence Testing. *Presented at the International Society of Clinical Biostatistics.*

[269] Patterson, S., Jones, B. (2002g) Simulation Assessment of Statistical Aspects of Bioequivalence in the Pharmaceutical Industry. *Pharmaceutical Statistics*, In Review.

[270] Patterson, S., Jones, B. (2002h) Statistical Aspects of Bioequivalence in the Pharmaceutical Industry. *Proceedings of the American Statistical Association Joint Statistical Meetings*, **2002**.

[271] Patterson, S., Jones, B. (2002i) Clinical Development Planning and the use of Pharmacokinetic Data in ICH-E5 Bridging Assessments. *GSK BDS Technical Report 2002-04.*

[272] Peace, K.E. (1986) Estimating the degree of equivalence and non-equivalence, an alternative to bioequivalence testing. *Proceedings of the American Statistical Association, Biopharmaceutical Chapter*, 63–69.

[273] Peace, K., ed. (1992) Biopharmaceutical Sequential Statistical Applications. Dekker.

[274] Phillips, K.F. (1990) Power of the two-one sided testing procedure in bioequivalence. *J Pharmacokinet Biopharm*, **18**, 137–144.

[275] Pitman, E.C.G. (1939) A note on normal correlation. *Biometrika*, **31**, 9–12.

[276] Pocock, S.J. (1977) Group Sequential Methods in the design and analysis of clinical trials. *Biometrics*, **64**, 191–199.

[277] Pound, N.J. (1999) Invited Presentation: Bioavailability and Bioequivalence: Update on Guidances from TPP. *AAPS International workshop on Individual Bioequivalence: Realities and Implementation.*

[278] Quiroz, J., Ting, N., Wei, G., Burdick, R. (2002) Alternative confidence intervals for the assessment of bioequivalence in four-period cross-over designs. *Statistics in Medicine*, **21**, 1825–1847.

[279] Racine-Poon, A., Grieve, A.P., Fluhler, H., Smith, A. F-M. (1986) Bayesian methods in practice, experiences in the pharmaceutical industry (with discussion). *Appl Stat*, **35**, 93–150.

[280] Racine-Poon, A., Grieve, A.P., Fluhler, H., Smith, A. F-M. (1987) A Two-Stage Procedure for Bioequivalence Studies. *Biometrics*, **43**, 847–856.

[281] Rao, C.R. (1973) Linear Statistical Inference and Its Applications, 2ed. John Wiley and Sons, NY.

[282] Reigner, B., Blesch, K. (2001) Estimating the starting dose for entry into humans: principles and practice. *Eur J Clin Pharmacol*, **57**, 835–845.

[283] Reisner, B., Guttman, I. (1992) Statistical inference for Pr(YY). *CommStA*, **21**, 559–569.

[284] Rescigno, A. (1992) Bioequivalence. *Pharmaceutical Research*, **9**, 925–928.

[285] Rescigno, A., Powers, J.D. (1998) AUC and Cmax are not sufficient to prove bioequivalence. *Pharmacological Research*, **37**, 93–95.

[286] Rheinstein, P.H. (1990) Therapeutic inequivalence. *Drug Safety 5*, **Suppl 1**, 114–119.

[287] Ring, A., Tothfalusi, L., Endrenyi, L., Weiss, M. (2000) Sensitivity of Empirical Metrics of Rate of Absorption in Bioequivalence Studies. *Pharmaceutical Research*, **17**, 583–588.

[288] Rocke, D.M. (1984) On testing for bioequivalence. *Biometrics*, **40**, 225–230.

[289] Rodda, B.E., Davis R.L. (1980) Determining the probability of an important difference in bioavailability. *Clin Pharmacol Ther*, **28**, 247–252.

[290] Rowland, M., Tozer, T.N.. (1980) *Clinical Pharmacokinetics Concepts and Applications.* Lea and Febidger, Philadelphia.

[291] Sarkar, S., Watts, S., Ohashi, O., Carroll, R., Uesaka, H., Mason, T., Rivera, C. (2002) Bridging Data Between Two Ethnic Populations. A New Application of Matched Case-Control Methodology *Drug Information Journal*, **36**, 349–56.

[292] Satterthwaite, F. (1941) Synthesis of Variance. *Psychometrika*, **6**, 309–316.

[293] Schall, R. (1995) Assessment of individual and population bioequivalence using the probability that bioavailabilities are similar. *Biometrics*, **51**, 615–626.

[294] Schall, R., Luus, H.G. (1993) On population and individual bioequivalence. *Statistics in Medicine*, **12**, 1109–1124.

[295] Schall, R., Williams, R.L. (1996) Towards a practical strategy for assessing individual bioequivalence. *J Pharmacokin Biopharm*, **24**, 133–149.

[296] Schwarz, G. (1978) Estimating the Dimension of a Model. *The Annals of Statistics*, **6(2)**, 461–464.

[297] Schuirmann, D.J. (1981) On hypothesis testing to determine if the mean of a normal distribution is contained in a known interval. *Biometrics*, **37**, 617.

[298] Schuirmann, D.J. (1987) A comparison of the two one sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokin Biopharm*, **15**, 657–680.

[299] Schuirmann, D.J. (1990) Design of Bioavailability and Bioequivalence Studies. *Drug Information Journal*, **24**, 315–323.

[300] Schulman, R.L. (1992) *Statistics in Plain English with Computer Applications.* Van Nostrand Reinhold, New York.

[301] Schumaker, R.C., Metzler, C.M. (1998) The phenytoin trial is a case study of individual bioequivalence. *Drug Information Journal*, **32**, 1063–1072.

[302] Searle, S.R. (1971) *Linear Models.* John Wiley and Sons, NY.

[303] Seber, G.A. (1977) *Linear Regression Analysis.* John Wiley and Sons, NY.

[304] Selwyn, M.R., Dempster, A.P., Hall, N.R. (1981) A Bayesian approach to bioequivalence for the 2x2 cross-over design. *Biometrics*, **37**, 11–21.

[305] Selwyn, M.R., Hall, N.R. (1984) On Bayesian Methods for Bioequivalence. *Biometrics*, **40**, 1103–1108.

[306] Senn, S. (1993) *Cross-over Trials in Clinical Research.* John Wiley and Sons, NY. (Second Edition, 2002).

[307] Senn, S. (1996) The AB-BA Cross-over: How to perform the two-stage analysis if you can't be persuaded you shouldn't. *in Hansen and De Ridder, Amicorum Roel van Strik*, **Erasmus U**, 93–100.

[308] Senn, S. (1997) *Statistical Issues in Drug Development.* John Wiley and Sons, NY.

[309] Senn, S. (1998) In the blood: proposed new requirements for registering generic drugs. *The Lancet*, **352**, 85–86.

[310] Senn, S., Ezzet, F. (1999) Clinical Cross-over Trials in Phase I. *Statistical Methods in Medical Research*, **8**, 263–78.

[311] Senn, S. (2000) Conference Proceedings: Challenging Statistical Issues in Clinical Trials. *Decisions and Bioequivalence.*

[312] Senn, S. (2001) Statistical Issues in Bioequivalence. *Statistics in Medicine*, **20**, 2785–2799.

[313] Senn, S. (2002) *Cross-over Trials in Clinical Research.* John Wiley and Sons, NY. (Second Edition)

[314] Serfling, R. (1980) *Approximation Theorems of Mathematical Statistics*, John Wiley and Sons, New York.

[315] Shah, V.P., Yacobi, A., Barr, W., Benet, L.Z., Breimer, D., Dobrinska, M.R., Endrenyi, L., Fairweather, W., Gillespie, W., Gonzalez, M.A., Hooper, J., Jackson, A., Lesko, L.J., Midha, K.K., Noonan, P.K., Patnaik, R., Williams, R.L. (1996) Workshop Report: Evaluation of Orally Administered Highly Variable Drugs and Drug Formulations. *Pharmaceutical Research*, **13**, 1590–1594.

[316] Shao, J., Tu, D. (1996) *The Jackknife and Bootstrap*, Springer, New York.

[317] Shao, J., Chow, S-C., Wang, B. (2000a) The bootstrap procedure in individual bioequivalence. *Statistics in Medicine*, **19**, 2741–2754.

[318] Shao, J., Kubler, J., Pigeot, I. (2000b) Consistency of the bootstrap procedure in individual bioequivalence. *Biometrika*, **87**, 573–585.

[319] Sheiner L.B., Beal, S.L., Sambol, N.C. (1989) Study Designs for Dose-Ranging. *Clin Pharm Ther*, **46**, 63–77.

[320] Sheiner, L.B. (1992) Bioequivalence revisited. *Statistics in Medicine*, **11**, 1777–1788.

[321] Sheiner, L.B. (1997) Learning versus confirming in clinical drug development. *Clinical Pharmacology and Therapeutics*, **61**, 275–291.

[322] Sheiner, L.B., Steimer, J-L. (2000) Pharmacokinetic-Pharmacodynamic Modeling in Drug Development. *Annu Rev Pharmacol Toxicol*, **40**, 67–95.

[323] Sheskin, D. (2000) *Handbook of Parametric and Nonparametric Statistical Procedures, 2ed*, Chapman and Hall, CRC.

[324] Singh, G.J-P., Adams, W.P., Lesko, L.J., Shah, V.P., Molzon, J.A., Williams, R.L., Pershing, L.K. (1999) Development of in vivo bioequivalence methodology for dermatologic corticosteroids based on pharmacodynamic modeling. *Clinical Pharmacology and Therapeutics*, **66**, 346–357.

[325] Snikeris, F., Tingey, H.B. (1994) A two step method for assessing bioequivalence. *Drug Information Journal*, **28**, 709–722.

[326] Spino, M., Tsang, Y.C., Pop, R. (2000) Dissolution and in vivo evidence of differences in reference products: impact on development of generic drugs. *European Journal of Drug Metabolism and Pharmacokinetics*, **25**, 18–24.

[327] Srinivasan, R, Langenberg, P. (1986) A Two-stage procedure with controlled error probabilities for testing equivalence. *Biom J*, **7**, 825–833.

[328] Steinjans, V.W., Diletti, E. (1983) Statistical analysis of bioavailability studies: parametric and nonparametric confidence interval. *Eur J Clin Pharm*, **24**, 127–136.

[329] Steinijans, V.W., Hauschke, V. (1990) Update on the statistical analysis of bioequivalence studies. *Int J Clin Pharmacol Ther Toxicol*, **28**, 105–110.

[330] Steinijans, V.W., Hartmann, M., Huber, R., Radtke, H.W. (1991) Lack of pharmacokinetic interaction as an equivalence problem. *Int J Clin Pharm Ther and Toxicol*, **29**, 323–328.

[331] Steinijans, V.W., Sauter, R., Hauschke, D., Elze, M. (1995) Metrics to Characterize Concentration-Time Profiles in Single and Multiple-Dose Bioequivalence Studies. *Drug Information Journal*, **29**, 981–987.

[332] Steinijans, V.W., Hauschke, D., Schall, R. (1995) International harmonization of regulatory requirements for average bioequivalence and current issues in individual bioequivalence. *Drug Information Journal*, **29**, 1055–1062.

[333] Steinjans, V.W., Diletti, E. (1997) Individual bioequivalence: a european perspective. *Journal of Biopharmaceutical Statistics*, **7**, 31–34.

[334] Strom, B.L. (1987) Generic Drug Substitution revisited. *NEJM*, **316**, 1456–1462.

[335] Stuart, A., Ord, K., Arnold, S. (1999) *Kendall's Advanced Theory of Statistics, Vol 2A - Classical Inference and the Linear Model, 6ed.*. Arnold, London.

[336] Swallow, W., Monahan, J. (1984) Monte carlo Comparison of ANOVA, MIVQUE, REML, and ML Estimators of Variance Components. *Technometrics*, **26**, 47–57.

[337] Tanida, N. (2002) Ethical Considerations for Clinical Trials in Asia. *Drug Information Journal*, **36**, 41–9.

[338] Ting, N., Burdick, R.K., Graybill, F.A., Jeyaratnam, S., Lu, T-F.C. (1990) Confidence intervals on linear combinations of variance components that are unrestricted in sign. *J Stat Comput Simul*, **35**, 135–143.

[339] Tozer, T.N., Bois, F.Y., Hauck, W.W., Chen, M.L., Williams, R.L. (1996) Absorption rate versus exposure: which is more useful for bioequivalence testing? *Pharmaceutical Research*, **13**, 453–456.

[340] Tozer, T.N., Hauck, W.W. (1997) Cmax/AUC, A Commentary. *Pharmaceutical Research*, **14**, 967–968.

[341] Tsang, Y.C., Opo, R., Gordon, P., Hems, J., Spino, M. (1996) High variability in drug pharmacokinetics complicates determination of bioequivalence: experience with verpamil. *Pharmaceutical Research*, **13**, 846–850.

[342] Vonesh, E.F., Chinchilli, V.M. (1997) *Linear and Nonlinear models for the analysis of repeated measurements*. Marcel Dekker, NY.

[343] Vuorinen, J. (1997) A practical approach for the assessment of bioequivalence under selected higher order cross-over designs. *Statistics in Medicine*, **16**, 2229–2243.

[344] Vuorinen, J., Turunen, J. (1996) A three step procedure for assessing bioequivalence in the general mixed linear framework. *Statistics in Medicine*, **15**, 2635–2655.

[345] Walpole, R.E., Myers, R.H., Myers, S.L. (1998) *Probability and Statistics for Engineers and Scientists, 6th ed.*, Prentice Hall, NJ.

[346] Wang, C.M. (1990) On the lower bound of confidence coefficients for a confidence interval on variance components. *Biometrics*, **46**, 187–192.

[347] Wang, W. (1997) Optimal unbiased tests for equivalence in intrasubject variability. *Journal of the American Statistical Association*, **88**, 939–946.

[348] Wang, W. (1999) On testing for individual bioequivalence. *Journal of the American Statistical Association*, **94**, 880–887.

[349] Wang, Y. (1999) Use of jacknife influence profiles in bioequivalence evaluations. *PSTT*, **2**, 152–159.

[350] Welleck, S. (1993) Basing the analysis of comparative bioavailability trials on an individualized statistical definition of equivalence. *Biometrical Journal*, **35**, 47–55.

[351] Welleck, S. (1997) A comment on so called individual criteria of bioequivalence. *Journal of Biopharmaceutical Statistics*, **7**, 17–21.

[352] Welleck, S. (2000) On a reasonable disaggregate criterion of population bioequivalence admitting of resampling-free testing procedures. *Statistics in Medicine*, **19**, 2755–2767.

[353] Welty, T.E., Pickering, P.R., Hale, B.C., Arazi, R. (1992) Loss of seizure control associated with generic substitution of carbamazepine. *Annals of Pharmacotherapy*, **26**, 775–777.

[354] Westlake, W.J. (1972) Use of confidence intervals in analysis of comparative bioavailability trials. *Journal ofPharmaceutical Sciences*, **61**, 1340–1341.

[355] Westlake, W.J. (1979) Statistical aspects of comparative bioavilability trials. *Biometrics*, **35**, 273–280.

[356] Westlake, W.J. (1986) Bioavailability and bioequivalence of pharmaceutical formulations. *Biopharmaceutical Statistics for Drug Development, K. Peace, ed.* Marcel Dekker, NY, 329–352.

[357] Wetherill, G., Glazebrook, K. (1986) Sequential Methods in Statistics. Chapman and Hall.

[358] Whitehead, J. (1996) Sequential designs for equivalence studies. *Statistics in Medicine*, **15**, 2703–2715.

[359] Williams, R.L., Adams, W., Chen, M-L., Hare, D., Hussain, A., Lesko, L., Patnaik, R., Shah, V., and FDA Biopharmaceutics Coordinating Committee. (2000a) Where are we now and where do we go next in terms of the scientific basis for regulation on bioavailability and bioequivalence? *European Journal of Drug Metabolism and Pharmacokinetics*, **25**, 7–12.

[360] Williams, R.L., Patnaik, R.N., Chen, M-L. (2000b) The basis for individual bioequivalence. *European Journal of Drug Metabolism and Pharmacokinetics*, **25**, 13–17.

[361] Wolfinger, R., Tobias, R., Sall, J. (1994) Computing Gaussian Likelihoods and Their Derivatives for General Linear Mixed Models. *Siam J Sci Comput*, **15**, 1294–1310.

[362] Yacobi, A., Masson, E., Moros, D., Ganes, D., Lapointe, C., Abolfathi, Z., LeBel, M., Golander, Y., Doepner, D., Blumberg, T., Cohen, Y., Levitt, B. (2000) Who needs individual bioequivalence studies for narrow therapeutic index drugs? A case for warfarin. *Journal of Clinical Pharmacology*, **40**, 826–835.

[363] Yasuhara, H. (1994) Which is more important in pharmacokinetics: inter-ethnic variability or intra-ethnic variability. *Proceedings of the Second International Conference on Harmonisation, 1993*, **D'Arcy, PF., Harron, D. eds**, Greystone Books, Ltd.

[364] Yamaoka, K., Nakagawa, T., Uno, T. (1978) Statistical Moments in Pharmacokinetics. *Journal of Pharmacokinetics and Biopharmaceutics*, **6**, 547–558.

[365] Yee, K.F. (1986) The calculation of probabilities in rejecting bioequivalence. *Biometrics*, **42**, 961–965.

[366] Yeh, K.C., Kwan, K.C. (1978) A Comparison of Numerical Integrating Algoritms of Trapezoidal, Legrange, And Spline Approximation. *J Pharmaco Biopharm*, **6**, 79–81.

[367] Zariffa, N.M–D., Patterson, S.D., Boyle, D., Hyneck, H. (1998) Invited Presentation: Case Studies, Comparison of the Current and Proposed Bioequivalence Criteria. *FDA/AAPS Workshop on Bioavailability and Bioequivalence.*

[368] Zariffa, N.M–D, Patterson, S.D. (1999) Invited presentation: A case review on the topic of subject by formulation interaction. *AAPS International workshop on Individual Bioequivalence: Realities and Implementation.*

[369] Zariffa, N.M–D, Patterson, S.D. (2000) Invited presentation: Learnings and Recommendations for Population and Individual Bioequivalence Based on the Analysis of Replicate Study Designs. *AAPS/FDA Workshop on Biopharmaceutics in the New Millenium: Regulatory Approaches to Bioavailability and Bioequivalence.*

[370] Zariffa, N.M–D., Patterson, S.D., Boyle, D., Hyneck, H. (2000) Case Studies, Practical Issues, and Observations on Population and Individual Bioequivalence. *Statistics in Medicine*, **19**, 2811–2820.

[371] Zariffa, N.M–D., Patterson, S.D. (2001) Population and Individual Bioequivalence: Lessons from Real Data and Simulation Studies. *J Clin Pharmacol*, **41**, 811–822.

8 Appendix: *SAS*© Simulation Code for Chapter 5-6

8.1 Two-Sequence Replicate Cross-over Design Simulation Code (Chapter 5)

```
*Macro for two sequence replicate design;
*Author: Scott Patterson;
*Date: 21JAN02;
*Sim macro outputs temp dataset forsim which is then analysed;

    %macro sim(mu1, /*mean T*/
mu2, /*mean R*/
sigbt, /*SD Between T*/
sigwt, /*SD Within T*/
sigbr, /*SD Between R*/
sigwr, /*SD Within R*/
f, /* sig2d=f(sig2wr)*/
i, /* Number of Sims */
j, /* Sample Size per sim */
seq, /* sequence effect */
p1, /* Period 1 effect */
p2, /* Period 2 effect */
p3, /* Period 3 effect */
p4); /* Period 4 effect */

    data bet1; keep i j t r; mu1=&mu1; mu2=&mu2; sigbt=&sigbt; sigwt=&sigwt; sigbr=&sigbr;
sigwr=&sigwr; f=&f;
    sig2bt=sigbt**2;sig2wt=sigwt**2;sig2br=sigbr**2;sig2wr=sigwr**2; sig2d=f*sig2wr;
    var1=sig2bt; var2=sig2br; rho=(sig2bt+sig2br-sig2d)/(2*sigbt*sigbr); c=sqrt(1-rho**2);
    do i = 1 to &i; do j = 1 to &j by 2; t = rannor(123); r = rho*t+c*rannor(123); t = mu1 +
sqrt(var1)*t; r = mu2 + sqrt(var2)*r; output bet1; end;end;
    run;

    data wit1a; keep i j x y; mu1=0; mu2=0; sigwt=&sigwt; sigwr=&sigwr;
    var1=(sigwt**2)/2; var2=(sigwr**2)/2;
    rho=0; c=sqrt(1-rho**2);
    do i = 1 to &i; do j = 1 to &j by 2; x = rannor(456); y = rho*x+c*rannor(456); x = mu1
+ sqrt(var1)*x; y = mu2 + sqrt(var2)*y; output wit1a; end;end; run; data all1a;merge bet1
wit1a;by i j; keep i j t r; t=t+x;r=r+y; run; proc sort;by i j;run;

    data wit1b; keep i j x y; mu1=0; mu2=0; sigwt=&sigwt; sigwr=&sigwr;
    var1=(sigwt**2)/2; var2=(sigwr**2)/2;
    rho=0; c=sqrt(1-rho**2);
    do i = 1 to &i; do j = 1 to &j by 2; x = rannor(789); y = rho*x+c*rannor(456); x = mu1
+ sqrt(var1)*x; y = mu2 + sqrt(var2)*y; output wit1b; end;end; run;

    data all1;merge all1a wit1b;by i j; keep i j t1 t2 r1 r2 t_bar r_bar; t1=t-x;t2=t+x;r1=r-
y;r2=r+y; t_bar=(t1+t2)/2;r_bar=(r1+r2)/2; sequence='RTRT'; keep i j sequence t1 t2 r1 r2
```

```
t_bar r_bar; run; proc sort;by i j;run;

    data bet2; keep i j t r; mu1=&mu1; mu2=&mu2; sigbt=&sigbt; sigwt=&sigwt; sigbr=&sigbr;
sigwr=&sigwr; f=&f; seq=&seq;
    sig2bt=sigbt**2;sig2wt=sigwt**2;sig2br=sigbr**2;sig2wr=sigwr**2; sig2d=f*sig2wr;
    var1=sig2bt; var2=sig2br; rho=(sig2bt+sig2br-sig2d)/(2*sigbt*sigbr); c=sqrt(1-rho**2);
    do i = 1 to &i; do j = 2 to &j by 2; t = rannor(1234); r = rho*t+c*rannor(1234); t = mu1
+ sqrt(var1)*t; r = mu2 + sqrt(var2)*r; *induce sequence effect; t = t+seq; r = r+seq; output
bet2; end;end;
    run;

    data wit2a; keep i j x y; mu1=0; mu2=0; sigwt=&sigwt; sigwr=&sigwr;
    var1=(sigwt**2)/2; var2=(sigwr**2)/2;
    rho=0; c=sqrt(1-rho**2);
    do i = 1 to &i; do j = 2 to &j by 2; x = rannor(4567); y = rho*x+c*rannor(4567); x = mu1
+ sqrt(var1)*x; y = mu2 + sqrt(var2)*y; output wit2a; end;end; run;

    data all2a;merge bet2 wit2a;by i j; keep i j t r; t=t+x;r=r+y; run; proc sort;by i j;run;
    data wit2b; keep i j x y; mu1=0; mu2=0; sigwt=&sigwt; sigwr=&sigwr;
    var1=(sigwt**2)/2; var2=(sigwr**2)/2;
    rho=0; c=sqrt(1-rho**2);
    do i = 1 to &i; do j = 2 to &j by 2; x = rannor(45678); y = rho*x+c*rannor(45678); x =
mu1 + sqrt(var1)*x; y = mu2 + sqrt(var2)*y; output wit2b; end;end; run;

    data all2;merge all2a wit2b;by i j; keep i j t1 t2 r1 r2 t_bar r_bar; t1=t-x;t2=t+x;r1=r-
y;r2=r+y; t_bar=(t1+t2)/2;r_bar=(r1+r2)/2; sequence='TRTR'; keep i j sequence t1 t2 r1 r2
t_bar r_bar; run; proc sort;by i j;run;

    data all;merge all1 all2;by i j; subject=j; *if subject=2 then t1=.; *if subject=2 then r1=.;
*if subject=1 then t1=.; *if subject=1 then t2=.; drop j; run;

    data p1;set all;period=1;
if sequence='RTRT' then do;ln_auc=r1;regimen='R';output;end;
if sequence='TRTR' then do;ln_auc=t1;regimen='T';output;end;
keep i subject sequence period regimen ln_auc;
run;
proc sort;by i subject sequence period regimen;run;
data p2;set all;period=2;
if sequence='RTRT' then do;ln_auc=t1;regimen='T';output;end;
if sequence='TRTR' then do;ln_auc=r1;regimen='R';output;end;
keep i subject sequence period regimen ln_auc;
run;
proc sort;by i subject sequence period regimen;run;
data p3;set all;period=3;
if sequence='RTRT' then do;ln_auc=r2;regimen='R';output;end;
if sequence='TRTR' then do;ln_auc=t2;regimen='T';output;end;
keep i subject sequence period regimen ln_auc;
run;
proc sort;by i subject sequence period regimen;run;
data p4;set all;period=4;
if sequence='RTRT' then do;ln_auc=t2;regimen='T';output;end;
if sequence='TRTR' then do;ln_auc=r2;regimen='R';output;end;
keep i subject sequence period regimen ln_auc;
run;
proc sort;by i subject sequence period regimen;run;

    data forsim;merge p1 p2 p3 p4;by i subject sequence period regimen;
per1eff=&p1;
per2eff=&p2;
per3eff=&p3;
```

```
per4eff=&p4;
if period=1 then ln_auc=ln_auc+per1eff;
if period=2 then ln_auc=ln_auc+per2eff;
if period=3 then ln_auc=ln_auc+per3eff;
if period=4 then ln_auc=ln_auc+per4eff;
lnmetric=ln_auc;
keep i subject sequence period regimen lnmetric; run;
    %mend sim;
```

8.2 Bootstrap Code (Chapter 5)

```
%macro bootstrp(indata=, /* dataset to sample from */ seq=, /* Sequence bootstrapping */
seed=12345, /* random number seed */ nrep=, /* # of bootstrap repetitions */ bootsamp=);
/* output dataset */ data test; set &indata;if sequence=&seq and sim=&sim; run;
data _null_; set test nobs=count; call symput('count',left(put(count,18.)));
if &nsamp¡=0 then call symput('nsamp',left(put(count,18.))); if &seed=0 then do; seed=time();
call symput('seed',left(put(seed,18.))); end; run;
data &bootsamp(label=seed=&seed); retain seed &seed; drop ijlkrst r seed; %do i=1 %to &nrep;
rep=&i; put "generating rep " rep "of &nrep"; do ijlkrst=1 to &nsamp; call ranuni(seed,r);
pointvar = ceil(r*&count); set test point=pointvar; output; end; %end; stop; run;
proc sort data=&bootsamp;by rep;run; %mend bootstrp;
```

8.3 Analysis Code (Chapter 5)

```
%macro anova(out,i1,i2);
    data dset;set forsim;if &i1¡=i¡=&i2; run;proc sort;by i subject;run;

    proc means data=dset noprint; by i subject; var lnmetric; output out=tot mean=mean; run;

    proc means data=tot noprint;by i; var mean; output out=n n=n; run;

    proc sort data=dset;by i sequence;run; proc means data=dset noprint; by i sequence; var ln-
metric; output out=seq mean=mean; run; proc means data=seq noprint;by i; var mean; output
out=p n=p; run;

    data n;merge n p;by i; run;

    **********Calc BLUE for diff in means using GLM;
    proc glm data=dset outstat=glmout noprint;by i; class regimen sequence subject period;
model lnmetric=sequence subject(sequence) period regimen regimen*subject(sequence)/SS1 SS3;
random sequence subject(sequence)/test;
lsmeans regimen/stderr cov e=regimen*subject(sequence) out=lsm noprint;
run;

    data tvalue;
set glmout;by i;
if _SOURCE_='REGIM*SUBJEC(SEQUEN)' and _TYPE_='SS3';
tval=tinv(0.95,df,0);
run;

    data A; set lsm (where=(regimen='T'));by i; mean_A=lsmean; var_A=cov2; cov_AB=cov1;
keep i mean_A var_A cov_AB; run;

    data B; set lsm (where=(regimen='R'));by i; mean_B=lsmean; var_B=cov1; keep i mean_B
var_B; run;

    data ci_AB;
merge tvalue A B;by i;
diff=mean_A-mean_B; ratio=exp(diff);
sed=sqrt(var_A+var_B-(2*cov_AB));
```

```
lnlow=diff-(tval*sed); low=exp(lnlow);
lnup=diff+(tval*sed); up=exp(lnup);
minus='A - B'; comp='A : B';
run;

    data point;set ci_ab;keep i diff ratio low up;run;

    **********Calc BLUE for diff in means using MIXED UN; proc sort data=dset;by i subject sequence period regimen;run;

    title 'UN FOR MACRO';run; proc mixed data=dset method=reml ITDETAILS DFBW CL=WALD ALPHA=0.1 scoring=50 maxiter=200 IC ASYCOV;by i; class regimen sequence subject period; model lnmetric=sequence period regimen/DDFM=SATTERTH; random regimen/type=UN subject=subject G; repeated/group=regimen subject=subject; lsmeans regimen sequence period; estimate 'T-R' regimen -1 1/CL ALPHA=0.1; ods output Estimates=BLUEUN CovParms=COVUN AsyCov=ascovun; run;

    data covdun(keep=i delta2 asigdel);set blueun;
delta2=ESTIMATE*ESTIMATE;
asigdel=STDERR*STDERR;
run;

    data covunBT(keep=i l_BT l_BTxw l_BTxBR l_BTxWT l_BTxWR); set ascovun; if COVPARM='UN(2,2)'; l_BT=COVP3; l_BTxw=COVP2; l_BTxBR=COVP1; l_BTxWT=COVP5; l_BTxWR=COVP4; run;

    data covunw(keep=i l_w l_BRxw l_wxWT l_wxWR); set ascovun; if COVPARM='UN(2,1)'; l_w=COVP2; l_BRxw=COVP1; l_wxWT=COVP5; l_wxWR=COVP4; run;

    data covunBR(keep=i l_BR l_BRxWT l_BRxWR); set ascovun; if COVPARM='UN(1,1)'; l_BR=COVP1; l_BRxWT=COVP5; l_BRxWR=COVP4; run;

    data covunWT(keep=i l_WT l_WTxWR); set ascovun; if COVPARM='Residual' and Row=5; l_WT=COVP5; l_WTxWR=COVP4; run;

    data covunWR(keep=i l_WR);
set ascovun;
if COVPARM='Residual' and Row=4;
l_WR=COVP4;
run;

    data blueUN(keep=i undiff unlow unup unratio unrlow unrup);set blueun;
undiff=ESTIMATE;unratio=exp(undiff);
unlow=LOWER;unrlow=exp(unlow);
unup=UPPER;unrup=exp(unup);
run;

    data bsigaun(keep=i bsdaun bsigaun);set covun;if substr(COVPARM,1,6)='UN(2,2';
bsigaun=ESTIMATE;
bsdaun=SQRT(ESTIMATE);
run;
data bsigbun(keep=i bsdbun bsigbun);set covun;if substr(COVPARM,1,6)='UN(1,1';
bsigbun=ESTIMATE;
bsdbun=SQRT(ESTIMATE);
run;
data wsigaun(keep=i wsdaun wsigaun);set covun;if substr(GROUP,9,1)='T';
wsigaun=ESTIMATE;
wsdaun=SQRT(ESTIMATE);
run;
data wsigbun(keep=i wsdbun wsigbun);set covun;if substr(GROUP,9,1)='R';
```

```
wsigbun=ESTIMATE;
wsdbun=SQRT(ESTIMATE);
run;
data covun(keep=i covun);set covun;if substr(COVPARM,1,6)='UN(2,1';
covun=ESTIMATE;
run;
```

data covun(keep=i bsigaun bsdaun bsigbun bsdbun covun wsigaun wsdaun wsigbun wsdbun sigdun rhoun); merge bsigaun bsigbun wsigaun wsigbun covun;by i; sigdun=bsigaun+bsigbun-(2*covun); rhoun=covun/(bsdaun*bsdbun); run;

data ascovun(keep=i v_ibeun v_cibeun f_ibeun f_cibeun ubibeun ubcibeun v_pbeun v_cpbeun f_pbeun f_cpbeun ubpbeun ubcpbeun); merge covdun covunBT covunw covunBR covunWT covunWR covun;by i; theta=(((log(1.25))**2)+0.05)/0.04; thetap=(((log(1.25))**2)+0.02)/0.04;

```
    v_ibeun=delta2+sigdun+wsigaun-((1+theta)*wsigbun);
v_cibeun=delta2+sigdun+wsigaun-wsigbun-(0.04*theta);
f_ibeun=(4*asigdel*delta2)+l_BT+l_BR+(4*l_w)+l_WT+((1+theta)*(1+theta)*l_WR)+
(2*l_BTxBR)-(4*l_BTxw)+(2*l_BTxWT)-(2*(1+theta)*l_BTxWR)-(4*l_BRxw)+(2*l_BRxWT)-

(2*(1+theta)*l_BRxWR)-(4*l_wxWT)+(4*(1+theta)*l_wxWR)-(2*(1+theta)*l_WTxWR);
f_cibeun=(4*asigdel*delta2)+
l_BT+l_BR+(4*l_w)+l_WT+l_WR+
(2*l_BTxBR)-(4*l_BTxw)+(2*l_BTxWT)-(2*l_BTxWR)-(4*l_BRxw)+(2*l_BRxWT)-
(2*l_BRxWR)-(4*l_wxWT)+(4*l_wxWR)-(2*l_WTxWR);
ubibeun=v_ibeun+((probit(0.95))*sqrt(f_ibeun));
ubcibeun=v_cibeun+((probit(0.95))*sqrt(f_cibeun));

    v_pbeun=delta2+bsigaun+wsigaun-((1+thetap)*(wsigbun+bsigbun));
v_cpbeun=delta2+bsigaun+wsigaun-(wsigbun+bsigbun)-(0.04*thetap);
f_pbeun=(4*asigdel*delta2)+
l_BT+l_WT+((1+thetap)*(1+thetap)*l_BR)+((1+thetap)*(1+thetap)*l_WR)+
(2*l_BTxWT)-(2*(1+thetap)*l_BTxBR)-(2*(1+thetap)*l_BTxWR)-
(2*(1+thetap)*l_BRxWT)-(2*(1+thetap)*l_WTxWR)+(2*(1+thetap)*(1+thetap)*l_BRxWR);
f_cpbeun=(4*asigdel*delta2)+l_BT+l_WT+l_BR+l_WR+
(2*l_BTxWT)-(2*l_BTxBR)-(2*l_BTxWR)-(2*l_BRxWT)-(2*l_WTxWR)+(2*l_BRxWR);
ubpbeun=v_pbeun+((probit(0.95))*sqrt(f_pbeun));
ubcpbeun=v_cpbeun+((probit(0.95))*sqrt(f_cpbeun));
run;
```

**********Calc BLUE for diff in means using MIXED CSH; proc sort data=dset;by i subject sequence period regimen;run;

title 'CSH FOR MACRO';run; proc mixed data=dset method=reml ITDETAILS DFBW CL=WALD ALPHA=0.1 scoring=50 maxiter=200 IC ASYCOV;

```
class regimen sequence subject period;by i;
model lnmetric=sequence period regimen/DDFM=SATTERTH;
random regimen/type=CSH subject=subject G;
repeated/group=regimen subject=subject;
lsmeans regimen sequence period;
estimate 'T-R' regimen -1 1/CL ALPHA=0.1;
ods output Estimates=BLUECSH CovParms=COVCSH AsyCov=ascovcsh;
run;
```

data covdcsh(keep=i delta2 asigdel);set bluecsh;

```
delta2=ESTIMATE*ESTIMATE;
asigdel=STDERR*STDERR;
run;
```

data covcshBT(keep=i l_BT l_BTxBR l_BTxr l_BTxWT l_BTxWR);

```
set ascovcsh;
if substr(COVPARM,1,5)='Var(2';
l_BT=COVP2;
l_BTxBR=COVP1;
l_BTxr=COVP3;
l_BTxWT=COVP5;
l_BTxWR=COVP4;
run;

    data covcshBR(keep=i l_BR l_BRxr l_BRxWT l_BRxWR);
set ascovcsh;
if substr(COVPARM,1,5)='Var(1';
l_BR=COVP1;
l_BRxr=COVP3;
l_BRxWT=COVP5;
l_BRxWR=COVP4;
run;

    data covcshr(keep=i l_r l_rxWT l_rxWR); set ascovcsh; if substr(COVPARM,1,3)='CSH';
l_r=COVP3; l_rxWT=COVP5; l_rxWR=COVP4; run;

    data covcshWT(keep=i l_WT l_WTxWR); set ascovcsh; if substr(COVPARM,1,3)='Res'
and ROW=5; l_WT=COVP5; l_WTxWR=COVP4; run;

    data covcshWR(keep=i l_WR); set ascovcsh; if substr(COVPARM,1,3)='Res' and ROW=4;
l_WR=COVP4; run;

    data bluecsh(keep=i cshdiff cshlow cshup cshratio cshrlow cshrup);
set bluecsh;
cshdiff=ESTIMATE;cshratio=exp(cshdiff);
cshlow=LOWER;cshrlow=exp(cshlow);
cshup=UPPER;cshrup=exp(cshup);
run;

    data bsigacsh(keep=i bsdacsh bsigacsh);
set covcsh;if substr(COVPARM,1,5)='Var(2';
bsigacsh=ESTIMATE;
bsdacsh=SQRT(ESTIMATE);
run;
data bsigbcsh(keep=i bsdbcsh bsigbcsh);
set covcsh;if substr(COVPARM,1,5)='Var(1';
bsigbcsh=ESTIMATE;
bsdbcsh=SQRT(ESTIMATE);
run;
data wsigacsh(keep=i wsdacsh wsigacsh);
set covcsh;if substr(GROUP,9,1)='T';
wsigacsh=ESTIMATE;
wsdacsh=SQRT(ESTIMATE);
run;
data wsigbcsh(keep=i wsdbcsh wsigbcsh);
set covcsh;if substr(GROUP,9,1)='R';
wsigbcsh=ESTIMATE;
wsdbcsh=SQRT(ESTIMATE);
run;
data rhocsh(keep=i rhocsh);
set covcsh;if substr(COVPARM,1,3)='CSH';
rhocsh=ESTIMATE;
run;

    data covcsh(keep=i bsigacsh bsdacsh bsigbcsh bsdbcsh covcsh wsigacsh wsdacsh wsigbcsh
```

```
wsdbcsh covcsh rhocsh sigdcsh);
merge bsigacsh bsigbcsh wsigacsh wsigbcsh rhocsh;by i;
covcsh=rhocsh*bsdacsh*bsdbcsh;
sigdcsh=bsigacsh+bsigbcsh-(2*covcsh);
run;

    data ascovcsh(keep=i v_pbecs v_cpbecs f_pbecs f_cpbecs ubpbecs ubcpbecs);
merge covdcsh covcshBT covcshBR covcshr covcshWT covcshWR covcsh;by i;
thetap=(((log(1.25))**2)+0.02)/0.04;

    v_pbecs=delta2+bsigacsh+wsigacsh-((1+thetap)*(wsigbcsh+bsigbcsh));
v_cpbecs=delta2+bsigacsh+wsigacsh-(wsigbcsh+bsigbcsh)-(0.04*thetap);
f_pbecs=(4*asigdel*delta2)+
l_BT+l_WT+((1+thetap)*(1+thetap)*l_BR)+((1+thetap)*(1+thetap)*l_WR)+
(2*l_BTxWT)-(2*(1+thetap)*l_BTxBR)-(2*(1+thetap)*l_BTxWR)-(2*(1+thetap)*l_BRxWT)-

(2*(1+thetap)*l_WTxWR)+(2*(1+thetap)*(1+thetap)*l_BRxWR);
f_cpbecs=(4*asigdel*delta2)+
l_BT+l_WT+l_BR+l_WR+
(2*l_BTxWT)-(2*l_BTxBR)-(2*l_BTxWR)-(2*l_BRxWT)-
(2*l_WTxWR)+(2*l_BRxWR);
ubpbecs=v_pbecs+((probit(0.95))*sqrt(f_pbecs));
ubcpbecs=v_cpbecs+((probit(0.95))*sqrt(f_cpbecs));
run;

    **********Calc BLUE for diff in means using MIXED FA0(2); proc sort data=dset;by i
subject sequence period regimen;run;

    title 'FA0(2) FOR MACRO';run;
proc mixed data=dset method=reml ITDETAILS DFBW CL=WALD ALPHA=0.1 scoring=50
maxiter=200 IC ASYCOV;by i; class regimen sequence subject period; model lnmetric=sequence
period regimen/DDFM=SATTERTH; random regimen/type=FA0(2) subject=subject G; re-
peated/group=regimen subject=subject; lsmeans regimen sequence period; estimate 'T-R' reg-
imen -1 1/CL ALPHA=0.1;
ods output Estimates=BLUEFA0 CovParms=COVFA0;
run;

    data bluefa0(keep=i fa0diff fa0low fa0up fa0ratio fa0rlow fa0rup);set bluefa0;
fa0diff=ESTIMATE;fa0ratio=exp(fa0diff);
fa0low=LOWER;fa0rlow=exp(fa0low);
fa0up=UPPER;fa0rup=exp(fa0up);
run;

    data bsigafa0(keep=i bsdafa0 bsigafa0);set covfa0;if substr(COVPARM,1,6)='FA(2,1';
bsigafa0=ESTIMATE*ESTIMATE;
bsdafa0=ESTIMATE;
run;
data bsigbfa0(keep=i bsdbfa0 bsigbfa0);set covfa0;if substr(COVPARM,1,6)='FA(1,1';
bsigbfa0=ESTIMATE*ESTIMATE;
bsdbfa0=ESTIMATE;
run;
data sigdfa0(keep=i sigdfa0);set covfa0;if substr(COVPARM,1,6)='FA(2,2';
sigdfa0=ESTIMATE*ESTIMATE;
run;
data wsigafa0(keep=i wsdafa0 wsigafa0);set covfa0;if substr(COVPARM,1,3)='Res' and sub-
str(GROUP,9,1)='T';
wsigafa0=ESTIMATE;
wsdafa0=SQRT(ESTIMATE);
run;
data wsigbfa0(keep=i wsdbfa0 wsigbfa0);set covfa0;if substr(COVPARM,1,3)='Res' and sub-
```

str(GROUP,9,1)='R';
wsigbfa0=ESTIMATE;
wsdbfa0=SQRT(ESTIMATE);
run;

data covfa0(keep=i bsigafa0 bsdafa0 bsigbfa0 bsdbfa0 sigdfa0 wsigafa0 wsdafa0 wsigbfa0 wsdbfa0); merge bsigafa0 bsigbfa0 wsigafa0 wsigbfa0 sigdfa0;by i; run;

**********Calc BLUE for diff in means using new model from 2.2; proc sort data=dset;by i subject sequence period regimen;run;

title 'RIRS FOR MACRO';run; proc mixed data=dset method=reml ITDETAILS DFBW CL=WALD ALPHA=0.1 scoring=50 maxiter=200 ASYCOV IC INFO;by i; class regimen sequence subject period; model lnmetric=sequence period regimen/S CHISQ DDFM=SATTERTH; random intercept regimen/type=SIMPLE subject=subject G V; repeated/group=regimen subject=subject R; estimate 'T-R' regimen -1 1/CL ALPHA=0.1; ods output Estimates=BLUERIS CovParms=COVRIS AsyCov=ascovris; run;

data covdris(keep=i delta2 asigdel);set blueris;
delta2=ESTIMATE*ESTIMATE;
asigdel=STDERR*STDERR;
run;

data covrisD(keep=i l_D l_DxWT l_DxWR);
set ascovris;
if COVPARM='REGIMEN';
l_D=COVP2;
l_DxWT=COVP4;
l_DxWR=COVP3;
run;

data covrisWT(keep=i l_WT l_WTxWR);
set ascovris;
if COVPARM='Residual' and Row=4;
l_WT=COVP4;
l_WTxWR=COVP3;
run;

data covrisWR(keep=i l_WR);
set ascovris;
if COVPARM='Residual' and Row=3;
l_WR=COVP3;
run;

data blueris(keep=i risdiff rislow risup risrat risrlow risrup);set blueris;
risdiff=ESTIMATE;risrat=exp(risdiff);
rislow=LOWER;risrlow=exp(rislow);
risup=UPPER;risrup=exp(risup);
run;

data bsigris(keep=i bsdris bsigris);
set covris;if substr(COVPARM,1,5)='Inter';
bsigris=ESTIMATE;
bsdris=SQRT(ESTIMATE);
run;
data sigdris(keep=i sigdris);
set covris;if substr(COVPARM,1,7)='REGIMEN';
sigdris=2*ESTIMATE;
run;
data wsigaris(keep=i wsdaris wsigaris);

```
set covris;if substr(GROUP,9,1)='T';
wsigaris=ESTIMATE;
wsdaris=SQRT(ESTIMATE);
run;
data wsigbris(keep=i wsdbris wsigbris);
set covris;if substr(GROUP,9,1)='R';
wsigbris=ESTIMATE;
wsdbris=SQRT(ESTIMATE);
run;
```

data covris(keep=i bsigris bsdris sigdris wsigaris wsdaris wsigbris wsdbris); merge bsigris wsigaris wsigbris sigdris;by i; run;

```
data ascovris(keep=i v_iberi v_ciberi f_iberi f_ciberi ubiberi ubciberi);
merge covdris covrisD covrisWT covrisWR covris;by i;
theta=(((log(1.25))**2)+0.05)/0.04;
v_iberi=delta2+sigdris+wsigaris-((1+theta)*wsigbris);
v_ciberi=delta2+sigdris+wsigaris-wsigbris-(0.04*theta);
f_iberi=(4*asigdel*delta2)+
(4*l_D)+l_WT+((1+theta)*(1+theta)*l_WR)+
(4*l_DxWT)-(4*(1+theta)*l_DxWR)-(2*(1+theta)*l_WTxWR);
f_ciberi=(4*asigdel*delta2)+
(4*l_D)+l_WT+l_WR+(4*l_DxWT)-(4*l_DxWR)-(2*l_WTxWR);
ubiberi=v_iberi+((probit(0.95))*sqrt(f_iberi));
ubciberi=v_ciberi+((probit(0.95))*sqrt(f_ciberi));
run;
```

*********Calc Variance of a mean obs(B + .5w);

```
proc sort data=dset;by i subject regimen sequence;run;
```

proc means data=dset noprint;by i subject regimen;id sequence; var lnmetric; output out=sigd mean=mean; run;

```
data sigda;set sigd;if regimen='T'; mean_a=mean;keep i sequence subject mean_a;run;proc
sort;by i sequence subject;run;
proc glm data=sigda outstat=outba noprint;by i; class sequence; model mean_a=sequence/SS1
SS3; run;
DATA bsiga(keep=i bsiga_w df_ta); set outba; if _SOURCE_='ERROR'; bsiga_w=ss/df; df_ta=df;
run;

data sigdb;set sigd;if regimen='R'; mean_b=mean;keep i sequence subject mean_b;run;proc
sort;by i sequence subject;run;
proc glm data=sigdb outstat=outbb noprint;by i; class sequence; model mean_b=sequence/SS1
SS3; run;

DATA bsigb(keep=i bsigb_w df_tb); set outbb; if _SOURCE_='ERROR'; bsigb_w=ss/df;
df_tb=df; run;
*********Calc sigma-i and diff in means acc to Hyslop approach;
proc sort data=dset;by i subject regimen sequence;run;
proc means data=dset noprint; by i subject regimen;id sequence; var lnmetric; output
out=sigi mean=mean; run;

data sigia;set sigi;if regimen='T'; mean_a=mean;keep i sequence subject mean_a;run; proc
sort;by i sequence subject;run;
data sigib;set sigi;if regimen='R'; mean_b=mean;keep i sequence subject mean_b;run; proc
sort;by i sequence subject;run;
data sigi;merge sigia sigib;by i sequence subject; diffmean=mean_a-mean_b; run;
proc sort data=sigi;by i sequence;run;
proc means data=sigi noprint;by i sequence; var diffmean; output out=dm n=n sum=sum;
run;
```

```
data dm(keep=i sequence ad_mean);set dm; ad_mean=sum/n; run;
proc means data=dm noprint;by i; var ad_mean; output out=dm n=n sum=sum; run;
data dm(keep=i d_mean);set dm; d_mean=sum/n; run;
data point;merge dm point;by i; if diff=. then diff=d_mean; if ratio=. then ratio=exp(d_mean);
run;
proc glm data=sigi outstat=glmouti NOPRINT;by i; class sequence;
model diffmean=sequence/SS1 SS3; run;
DATA SIGI; set glmouti; if _source_='ERROR'; df_i=df; sigi=ss/df; keep i df_i sigi; run;
**********Calculate within-subject variances;
    proc sort data=dset;by i regimen sequence subject period;run;
    proc glm data=dset outstat=glmoutw noprint; class sequence subject period; by i regimen;
model lnmetric=sequence subject(sequence) period/SS1 SS3; run;
DATA SIGWA; set glmoutw; if _source_='ERROR' and regimen='T'; df_wa=df; sigwa=ss/df;
keep i df_wa sigwa; run;
DATA SIGWB; set glmoutw; if _source_='ERROR' and regimen='R'; df_wb=df; sigwb=ss/df;
keep i df_wb sigwb; run;
**********Merge & Calculate 90% CI;
    data final; merge point blueun bluefa0 bluecsh blueris n sigwa sigwb sigi bsiga bsigb covun
covcsh covfa0 covris ascovun ascovcsh ascovris;by i;
sigd=sigi-(0.5*(sigwa+sigwb));
if low=. then low=exp(diff-((tinv(0.95,df_i,0))*(SQRT(sigi/n))));
if up=. then up=exp(diff+((tinv(0.95,df_i,0))*(SQRT(sigi/n))));

    keep i diff ratio low up n p df_wa sigwa df_wb sigwb df_i sigi sigd df_ta df_tb bsiga_w bsigb_w
risdiff rislow risup risrat risrlow risrup undiff unlow unup unratio unrlow unrup fa0diff fa0low
fa0up fa0ratio fa0rlow fa0rup cshdiff cshlow cshup cshratio cshrlow cshrup bsigaun bsdaun bsig-
bun bsdbun covun wsigaun wsdaun wsigbun wsdbun sigdun rhoun bsigafa0 bsdafa0 bsigbfa0
bsdbfa0 sigdfa0 wsigafa0 wsdafa0 wsigbfa0 wsdbfa0 bsigacsh bsdacsh bsigbcsh bsdbcsh covcsh
wsigacsh wsdacsh wsigbcsh wsdbcsh sigdcsh rhocsh bsigris bsdris sigdris wsigaris wsdaris wsig-
bris wsdbris v_ibeun v_cibeun f_ibeun f_cibeun ubibeun ubcibeun v_pbeun v_cpbeun f_pbeun
f_cpbeun ubpbeun ubcpbeun v_pbecs v_cpbecs f_pbecs f_cpbecs ubpbecs ubcpbecs v_iberi v_ciberi
f_iberi f_ciberi ubiberi ubciberi ; run;
data final;set final;
    sigmad=0; if sigd¿0 then sigmad=sqrt(sigd);
lowdci=diff-((tinv(0.95,df_i,0))*(SQRT(sigi/n)));
highdci=diff+((tinv(0.95,df_i,0))*(SQRT(sigi/n)));
HD=((abs(diff))+((tinv(0.95,df_i,0))*(SQRT(sigi/n))))**2;
e_dcil=exp(lowdci);e_dcih=exp(highdci);

    HI=(df_i*sigi)/(cinv(0.05,df_i,0));

    sdwa=sqrt(sigwa);sdwb=sqrt(sigwb);
highwa=(df_wa*sigwa)/(cinv(0.05,df_wa,0));
highwb=(df_wb*sigwb)/(cinv(0.05,df_wb,0));

    run;
data &out;set final;
    ED=diff*diff;
if ED=. then ED=(diff*diff);
if HD=. then HD=((abs(diff))+((tinv(0.95,df_i,0))*(SQRT(sigi/n))))**2;
UD=(HD-ED)*(HD-ED);

    UI=(HI-sigi)*(HI-sigi);

    EA=0.5*sigwa;
HA=0.5*highwa;
UA=(HA-EA)*(HA-EA);

    theta=(((log(1.25))**2)+0.05)/0.04;
```

```
    EB=(-1.5-theta)*sigwb;
HB=(df_wb*EB)/(cinv(0.95,df_wb,0));
UB=(HB-EB)*(HB-EB);

    EBC=(-1.5)*sigwb;
HBC=(df_wb*EBC)/(cinv(0.95,df_wb,0));
UBC=(HBC-EBC)*(HBC-EBC);

    nu1=ED+sigi+EA+EB;
nu2=ED+sigi+EA+EBC-(theta*0.04);

    IBETRUE=nu1+(SQRT(UD+UI+UA+UB));
IBECTRUE=nu2+(SQRT(UD+UI+UA+UBC));

IBECRIT=theta+nu1+(SQRT(UD+UI+UA+UB));
IBECONST=theta+nu2+(SQRT(UD+UI+UA+UBC));

    **********Relative contribution of spread to mean ratio;
    RELD=(sqrt(UD))/ED;
RELI=(sqrt(UI))/sigi;
RELA=(sqrt(UA))/EA;
RELB=(sqrt(UB))/EB;

    **********PopBE Assessment;
    thetap=(((log(1.25))**2)+0.02)/0.04;

    totsiga=bsiga_w+(0.5*sigwa);
totsigb=bsigb_w+(0.5*sigwb);
sdtotb=sqrt(totsigb);

    /* ETT=totsiga;
HTT=(df_ta*ETT)/(cinv(0.05,df_ta,0));
UTT=(HTT-ETT)*(HTT-ETT);

    ETR=(-1-thetap)*totsigb;
HTR=(df_tb*ETR)/(cinv(0.95,df_tb,0));
UTR=(HTR-ETR)*(HTR-ETR);

    ETRS=(-1)*totsigb;
HTRS=(df_tb*ETRS)/(cinv(0.95,df_tb,0));
UTRS=(HTRS-ETRS)*(HTRS-ETRS);
*/
    EA_W=BSIGA_W;
HA_W=(DF_TA*EA_W)/(cinv(0.05,df_ta,0));
UA_W=(HA_W-EA_W)**2;

    EWA=0.5*SIGWA;
HWA=(DF_WA*EWA)/(cinv(0.05,df_wa,0));
UWA=(HWA-EWA)**2;
EB_W=(-1-thetap)*BSIGB_W;
HB_W=(DF_TB*EB_W)/(cinv(0.95,df_tb,0));
UB_W=(HB_W-EB_W)**2;

    EWB=(-1-thetap)*0.5*SIGWB;
HWB=(DF_WB*EWB)/(cinv(0.95,df_wb,0));
UWB=(HWB-EWB)**2;

    EB_WC=(-1)*BSIGB_W;
HB_WC=(DF_TB*EB_WC)/(cinv(0.95,df_tb,0));
```

```
UB_WC=(HB_WC-EB_WC)**2;

    EWBC=(-1)*0.5*SIGWB;
HWBC=(DF_WB*EWBC)/(cinv(0.95,df_wb,0));
UWBC=(HWBC-EWBC)**2;

    P1=ED+EA_W+EWA+EB_W+EWB;
P2=ED+EA_W+EWA+EB_WC+EWBC-(thetap*0.04);

    POPTRUE=P1+(SQRT(UD+UA_W+UWA+UB_W+UWB));
POPTRUEC=P2+(SQRT(UD+UA_W+UWA+UB_WC+UWBC));

    POPCRIT=thetap+P1+(SQRT(UD+UA_W+UWA+UB_W+UWB));
POPCONST=thetap+P2+(SQRT(UD+UA_W+UWA+UB_WC+UWBC));
```

```
    **********Variance components for section 2.5;
    tsigaun=bsigaun+wsigaun;
tsigbun=bsigbun+wsigbun;
tsigacsh=bsigacsh+wsigacsh;
tsigbcsh=bsigbcsh+wsigbcsh;
tsigafa0=bsigafa0+wsigafa0;
tsigbfa0=bsigbfa0+wsigbfa0;
tsigaris=bsigris+wsigaris;
tsigbris=bsigris+wsigbris;

    **********SE for ABE Assessment;
    VARMOM=sigd+((sigwa+sigwb)/2);
VARUN=sigdun+((wsigaun+wsigbun)/2);
VARCSH=sigdcsh+((wsigacsh+wsigbcsh)/2);
VARFA0=sigdfa0+((wsigafa0+wsigbfa0)/2);
VARRIS=sigdris+((wsigaris+wsigbris)/2);
```

```
    ****Differences in means;
    momdiff=log(ratio);
undiff=log(unratio);
csdiff=log(cshratio);
fadiff=log(fa0ratio);
ridiff=log(risrat);

    **********keep statement;
keep i n ratio low up unratio unrlow unrup fa0ratio fa0rlow fa0rup cshratio cshrlow cshrup risrat risrlow risrup df_wa sigwa sdwa df_wb sigwb sdwb df_i sigi sigd sigmad lsigmad usigmad nu1 nu2 ibecrit ibeconst p1 p2 POPCRIT POPCONST POPTRUE POPTRUEC df_ta df_tb bsiga_w bsigb_w totsiga totsigb sdtotb totratio t_ratiol t_ratiou wratio w_ratiol w_ratiou t5ratiol t5ratiou w5ratiol w5ratiou kl IBETRUE IBECTRUE RELD RELI RELA RELB UD UI UA UB UBC bsigaun bsigbun wsigaun wsigbun sigdun covun tsigaun tsigbun bsigafa0 bsigbfa0 sigdfa0 wsigafa0 wsigbfa0 tsigafa0 tsigbfa0 bsigacsh bsigbcsh wsigacsh wsigbcsh sigdcsh rhocsh tsigacsh tsigbcsh bsigris wsigaris wsigbris sigdris tsigaris tsigbris VARMOM VARUN VARCSH VARFA0 VARRIS v_ibeun v_cibeun f_ibeun f_cibeun ubibeun ubcibeun v_pbeun v_cpbeun f_pbeun f_cpbeun ubpbeun ubcpbeun v_pbecs v_cpbecs f_pbecs f_cpbecs ubpbecs ubcpbecs v_iberi v_ciberi f_iberi f_ciberi ubiberi ubciberi momdiff undiff csdiff fadiff ridiff ; run;
%mend;
```

8.4 Parallel Design Simulation Code (Chapter 6)

```
%macro sim(mu1, /*mean T*/ mu2, /*mean R*/ sigt, /*SD T*/ sigr, /*SD R*/ i, /* Number of Sims */ jt, /* Sample Size per sim population t*/ jr) /* Sample Size per sim population r*/
;
data t; keep i j race lnmetric; race='T'; mu1=&mu1; sigt=&sigt; sig2t=sigt**2; var1=sig2t; do
```

```
i = 1 to &i; do j = 1 to &jt; t = rannor(123); lnmetric = mu1 + sqrt(var1)*t; output t; end;end;
run;
data r; keep i j race lnmetric; race='R'; mu2=&mu2; sigr=&sigr; sig2r=sigr**2; var2=sig2r; do
i = 1 to &i; do j = 1 to &jr; r = rannor(456); lnmetric = mu2 + sqrt(var2)*r; output r; end;end;
run;
data forsim;set t r;by i j; subject=j; drop j; run;
    proc sort;by i subject race;run;
    %mend sim;
```

8.5 Bootstrap Code (Chapter 6)

```
%macro bootstrp(indata=, /* dataset to sample from */ i=, /* Sim run */ race=, /* Race */
seed=12345, /* random number seed */ nrep=, /* # of bootstrap repetitions */ nsamp=0, /* #
in bootstrap sample */ bootsamp=); /* output dataset */ data test; set &indata;if race=&race
and i=&i; run;
data _null_; set test nobs=count; call symput('count',left(put(count,18.))); if &nsamp¡=0 then
call symput('nsamp',left(put(count,18.)));
if &seed=0 then do; seed=time(); call symput('seed',left(put(seed,18.))); end; run;
data &bootsamp(label=seed=&seed); retain seed &seed; drop ijlkrst r seed; %do b=1 %to
&nrep; rep=&b; put "generating rep " rep "of &nrep"; do ijlkrst=1 to &nsamp; call ra-
nuni(seed,r); pointvar = ceil(r*&count); set test point=pointvar; output; end; %end; stop;
run; %mend bootstrp;
```

8.6 Analysis Code (Chapter 6)

```
%macro boot_mix(in);
    data boot_it1;set &in;by rep;if rep¡=500;run; data boot_it2;set &in;by rep;if 500¡rep¡=1000;run;
proc sort;by rep race;run;
ods listing close; proc mixed data=boot_it1 method=reml scoring=50 maxiter=200 itdetails
CL alpha=0.1; by rep; class race; model lnmetric=race/S DDFM=KENWARDROGER; re-
peated /group=race; estimate 'T - R' race -1 1/alpha=0.10; ods output Estimates=delbt1 Cov-
Parms=COVBT1; run;
proc mixed data=boot_it2 method=reml scoring=50 maxiter=200 itdetails CL alpha=0.1; by
rep; class race; model lnmetric=race/S DDFM=KENWARDROGER; repeated /group=race;
estimate 'T - R' race -1 1/alpha=0.10; ods output Estimates=delbt2 CovParms=COVBT2;
run;
ods listing;
    data delbt(keep=rep delta);set delbt1 delbt2 ;by rep; delta=ESTIMATE; run;
data sig2_tbt(keep=rep sig2_t); set covbt1 covbt2 ;
by rep;if substr(Group,6,1)='T'; sig2_t=ESTIMATE; run;
data sig2_rbt(keep=rep sig2_r);
set covbt1 covbt2 ;by rep;if substr(Group,6,1)='R';
sig2_r=ESTIMATE; run;

    data boot_out;merge delbt sig2_tbt sig2_rbt;
fda=((delta*delta)+sig2_t-sig2_r)/sig2_r;
kld=0.5*((delta*delta)+sig2_t+sig2_r)*((1/sig2_t)+(1/sig2_r))-2;
t_r=sig2_t/sig2_r;
run;

    proc univariate data=boot_out plot; var delta sig2_t sig2_r t_r fda kld; output out=&in
p5=delta5 sig2_t5 sig2_r5 t_r5 fda5 kld5
mean=deltam sig2_tm sig2_rm t_rm fdam kldm p95=delta95 sig2_t95 sig2_r95 t_r95 fda95 kld95
; run;

    %mend boot_mix;
```

9 Appendix: Summary Tables

Table 32: Sample Size, Sequences, and Number of Missing Observations for Data Sets A through ZF

Data Set	Sample Size	Sequences	Number of Missing Observations
A	30	TTRR,RRTT,RTTR,TRRT	3 AUC, 3 Cmax
B	74	TRRT,RTTR	7 AUC, 0 Cmax
C1	30	TTRR,RRTT,RTTR,TRRT	2 AUC, 2 Cmax
C2	29	TTRR,RRTT,RTTR,TRRT	NA
D	32	TTRR,RRTT,RTTR,TRRT	NA
E	17	RTTR,TRRT	1 AUC, 1 Cmax
F	12	RTRT,TRTR	NA
G	16	TTRR,RRTT,RTTR,TRRT	NA
H	20	TTRR,RRTT,RTTR,TRRT,TRTR	NA
I1	24	RTTR,TRRT	NA
I2	24	RTTR,TRRT	NA
J	24	TTRR,RRTT,RTTR,TRRT	NA
K1	36	TTRR,RRTT,RTTR,TRRT	NA
K2	36	TTRR,RRTT,RTTR,TRRT	NA
K3	36	TTRR,RRTT,RTTR,TRRT	NA
L1	38	RTTR,TRRT	3 AUC, 3 Cmax
L2	38	RTTR,TRRT	3 AUC, 3 Cmax
M	20	TTRR,RRTT,RTTR,TRRT	NA
N1	28	TTRR,RRTT,RTTR,TRRT	NA
N2	28	TTRR,RRTT,RTTR,TRRT	1 AUC, 0 Cmax
O1	24	TTRR,RRTT,RTTR,TRRT	NA
O2	24	TTRR,RRTT,RTTR,TRRT	NA
P	24	RTTR,TRRT	7 AUC, 7 Cmax
Q1	29	RTTR,TRRT	5 AUC, 5 Cmax
Q2	33	RTTR,TRRT	6 AUC, 6 Cmax
R	75	RTTR,TRRT	18 AUC, 12 Cmax
S	95	RTTR,TRRT	14 AUC, 4 Cmax
T	96	RTTR,TRRT	2 AUC, 2 Cmax
U	40	TTRR,RRTT,RTTR,TRRT	5 AUC, 4 Cmax
V	25	TTRR,RRTT,RTTR,TRRT	2 AUC, 2 Cmax
W1	36	TTRR,RRTT,RTTR,TRRT	NA
W2	36	TTRR,RRTT,RTTR,TRRT	NA
W3	36	TTRR,RRTT,RTTR,TRRT	NA
W4	36	TTRR,RRTT,RTTR,TRRT	NA
W5	36	TTRR,RRTT,RTTR,TRRT	NA
W6	36	TTRR,RRTT,RTTR,TRRT	NA
X	20	TTRR,RRTT,RTTR,TRRT	NA
Y	20	TTRR,RRTT,RTTR,TRRT,TRTR	12 AUC, 12 Cmax
ZA	22	TTRR,RTRT,TRTR,RRTT	NA
ZB	19	TTRR,RTRT,TRTR,RRTT	NA
ZC1	43	TTRR,RTTR,TRRT,RRTT	NA
ZC2	43	TTRR,RTTR,TRRT,RRTT	NA
ZC3	43	TTRR,RTTR,TRRT,RRTT	3 AUC, 0 Cmax
ZD1	38	TRRT,RTTR	NA
ZD2	38	TRRT,RTTR	NA
ZD3	38	TRRT,RTTR	NA
ZD4	38	TRRT,RTTR	6 AUC, 0 Cmax
ZE1	37	TRRT,RTTR	NA
ZE2	37	TRRT,RTTR	0 AUC, 1 Cmax
ZE3	37	TRRT,RTTR	NA
ZF	54	RTRT,TRTR	7 AUC, 7 Cmax

R=Reference, T=Test, NA=None Missing

Table 33: Gender, Ethnicity, Age, Weight, and Height for Data Sets A through ZF

Data Set	Gender	Ethnicity	Age	Weight	Height
A	0F,30M	9B,3O,18W	28±6	79±11	178±8
B	22F,52M	74W	33±8	75±12	176±8
C1	18F,12M	1B,29W	36±10	72±13	171±10
C2	17F,12M	1B,28W	36±10	72±13	171±9
D	32M	32W	28±6	76±9	178±8
E	17M	17W	37±8	81±11	177±6
F	6F,6M	12W	26±3	79±4	164±31
G	16M	16W	27±5	78±6	182±5
H	20M	1O,19W	47±9	93±14	183±8
I1	9F,15M	24W	37±10	76±14	175±11
I2	9F,15M	24W	37±10	76±14	175±11
J	7F,17M	24W	36±10	76±11	177±8
K1	36M	23B,13W	34±6	76±11	179±6
K2	36M	23B,13W	34±6	76±11	179±6
K3	36M	23B,13W	34±6	76±11	179±6
L1	13F,25M	38W	37±12	77±12	178±10
L2	13F,25M	38W	37±12	77±12	178±10
M	20M	20W	28±5	78±7	179±5
N1	10F,18M	28W	31±7	75±12	177±8
N2	10F,18M	28W	31±7	75±12	177±8
O1	8F,16M	1B,1O,22W	28±5	73±12	175±10
O2	8F,16M	1B,1O,22W	28±5	73±12	175±10
P	NA	NA	NA	NA	NA
Q1	19F,10M	2B,1O,26W	36±9	67±10	170±8
Q2	22F,11M	2B,1O,30W	35±9	67±10	170±8
R	38F,37M	1B,74W	31±8	69±10	174±9
S	38F,57M	95W	35±7	74±12	175±9
T	46F,50M	1B,2O,93W	33±9	69±10	172±8
U-ZE3	NA	NA	NA	NA	NA
ZF	15F,39M	28B,2O,24W	34±10	76±11	174±10

NA: Not Available
Gender: F=Female, M=Male
Ethnicity: B=Black, O=Other,W=White
Age(yrs), Weight(kg), and Height(cm) expressed as Mean±SD

Table 34: REML Model Discrimination using the Akaike Information Criterion (AIC) and the Schwarz Bayesian Criterion (SBC) for AUC in Data Sets A through ZF

Data Set	AIC UN	AIC CSH	AIC FA0(2)	AIC RIS	SBC UN	SBC CSH	SBC FA0(2)	SBC RIS
A	-51.5	-51.7	-51.7	-50.9	-58.3	-58.4	-58.4	-56.3
B	-297	-297	-297	-296	-306	-306	-306	-303
C1	-0.8	-1.2	-1.2	-0.3	-7.6	-8	-8	-5.7
C2	-39.6	-39.6	-39.6	-38.8	-46.3	-46.3	-46.3	-44.2
D	-16.3	-16.3	-16.3	-15.3	-23.3	-23.3	-23.3	-20.9
E	19.9	19.9	19.8	20.8	14.6	14.6	14.6	16.6
F	-20.2	-22.3	-22.3	-21.5	-24.5	-26.6	-26.6	-25
G	-17.5	-17.5	-17.5	-17.8	-22.6	-22.6	-22.6	-21.8
H	8.7	8.7	8.7	9.5	3.1	3.1	3.1	5
I1	-28.9	-28.9	-28.9	-28.6	-35.2	-35.2	-35.2	-33.6
I2	-69	-69	-69	-70.6	-75.3	-75.3	-75.3	-75.6
J	-41	-41	-41	-40.4	-47.2	-47.2	-47.2	-45.4
K1	-52.7	-53	-53	-52.2	-60	-60.3	-60.3	-58
K2	37.9	37.9	37.9	37	30.6	30.6	30.6	31.2
K3	4.8	1.2	1.2	2.1	-2.5	-6.1	-6.1	-3.7
L1	-51.2	-51.2	-51.2	-50.3	-58.6	-58.6	-58.6	-56.2
L2	-101	-103	-103	-102	-109	-110	-110	-108
M	-23.9	-23.9	-23.9	-23.2	-29.6	-29.6	-29.6	-27.7
N1	29.5	29.2	29.2	30	22.9	22.6	22.6	24.7
N2	-34.4	-34.4	-34.4	-33.8	-41	-41	-41	-39.1
O1	-7	-7	-7	-6.4	-13.2	-13.2	-13.2	-11.3
O2	-40	-40	-40	-40	-46.2	-46.2	-46.2	-45
P	25.7	25.2	25.2	26.2	19.7	19.2	19.2	21.4
Q1	27.8	27.8	27.8	26.4	21.1	21.1	21.1	21.1
Q2	-28.1	-28.1	-28.1	-27.1	-35	-35	-35	-32.6
R	-317	-317	-317	-316	-326	-326	-327	-324

UN=Unstructured
CSH=Heterscedastic Compound Symmetry
FA0(2)=First order autoregressive
RIS=Random Intercept and Slope
See Section 2.2 for details

Table 34: REML Model Discrimination using the Akaike Information Criterion (AIC) and the Schwarz Bayesian Criterion (SBC) for AUC in Data Sets A through ZF

Data Set	AIC UN	AIC CSH	AIC FA0(2)	AIC RIS	SBC UN	SBC CSH	SBC FA0(2)	SBC RIS
S	-439	-439	-439	-439	-449	-449	-449	-447
T	-462	-462	-462	-462	-472	-472	-472	-470
U	-125	-127	-127	-126	-133	-134	-134	-132
V	-63.6	-63.6	-63.6	-62.7	-69.9	-69.9	-69.9	-67.7
W1	-5.6	-6	-6	-5	-12.8	-13.2	-13.2	-10.8
W2	-15	-15.2	-15.2	-14.2	-22.2	-22.4	-22.4	-19.9
W3	-16.4	-16.6	-16.6	-16.4	-23.6	-23.8	-23.8	-22.1
W4	-55.5	-55.5	-55.5	-55.1	-62.7	-62.7	-62.7	-60.9
W5	-32	-32	-32	-31.1	-39.3	-39.3	-39.3	-36.9
W6	-87.7	-87.7	-87.7	-86.7	-94.9	-94.9	-94.9	-92.5
X	-16.8	-17.8	-17.8	-17.3	-22.4	-23.5	-23.5	-21.9
Y	-11.6	-11.6	-11.6	-11.2	-17.2	-17.2	-17.2	-15.8
ZA	-39.7	-39.7	-39.7	-38.7	-45.6	-45.6	-45.6	-43.4
ZB	-20.9	-20.9	-20.9	-19.9	-26.4	-26.4	-26.4	-24.4
ZC1	-83	-83	-83	-83.2	-90.8	-90.8	-90.8	-89.3
ZC2	-7.2	-7.2	-7.2	-8.3	-14.9	-14.9	-14.9	-14.5
ZC3	-18.4	-18.4	-18.4	-20.1	-26.1	-26.1	-26.1	-26.3
ZD1	-139	-148	-148	-149	-146	-156	-156	-155
ZD2	38.3	33.9	33.9	33.1	30.9	26.4	26.4	27.1
ZD3	33.4	33.3	33.3	32.9	25.9	25.9	25.9	26.9
ZD4	21.4	21.4	21.4	21.9	14.1	14.1	14.1	16.1
ZE1	-98.3	-98.3	-98.3	-97.4	-106	-106	-106	-103
ZE2	5.4	5.4	5.4	5.6	-2	-2	-2	-0.3
ZE3	-5.4	-5.4	-5.4	-7.2	-12.8	-12.8	-12.8	-13.1
ZF	-128	-128	-128	-127	-136	-136	-136	-134

UN=Unstructured
CSH=Heterscedastic Compound Symmetry
FA0(2)=First order autoregressive
RIS=Random Intercept and Slope
See Section 2.2 for details

Table 35: Residual *log*−likelihoods for AUC in Data Sets A through ZF

Data Set	UN	FA0(2)	CSH	RIS
A	-46.543	-46.696	-46.696	-46.939
B	-291.731	-292.135	-292.135	-292.159
C1	4.197	3.771	3.771	3.676
C2	-34.624	-34.624	-34.624	-34.845
D	-11.305	-11.305	-11.305	-11.306
E	24.859	24.845	24.859	24.838
F	-15.168	-17.257	-17.257	-17.532
G	-12.539	-12.539	-12.539	-13.77
H	13.731	13.731	13.731	13.543
I1	-23.901	-23.905	-23.905	-24.615
I2	-64.009	-64.009	-64.009	-66.576
J	-35.97	-35.97	-35.97	-36.421
K1	-47.724	-48.014	-48.014	-48.166

UN=Unstructured
CSH=Heterscedastic Compound Symmetry
FA0(2)=First order autoregressive
RIS=Random Intercept and Slope
See Section 2.2 for details

Table 35: Residual *log*−likelihoods for AUC in Data Sets A through ZF

Data Set	UN	FA0(2)	CSH	RIS
K2	42.911	42.911	42.911	41.006
K3	9.802	6.217	6.217	6.137
L1	-46.156	-46.183	-46.183	-46.276
L2	-96.364	-97.696	-97.696	-97.696
M	-18.918	-18.918	-18.918	-19.151
N1	34.516	34.238	34.238	33.961
N2	-29.412	-29.412	-29.412	-29.794
O1	-2.013	-2.013	-2.013	-2.35
O2	-34.978	-34.978	-34.978	-36.019
P	30.706	30.212	30.212	30.208
Q1	32.754	32.754	32.754	30.395
Q2	-23.063	-23.063	-23.063	-23.072
R	-312.39	-312.457	-312.39	-312.398
S	-434.036	-434.066	-434.066	-435.099
T	-457.466	-457.466	-457.466	-457.782
U	-120.37	-121.573	-121.573	-122.312
V	-58.627	-58.627	-58.627	-58.709
W1	-0.564	-1.031	-1.031	-1.041
W2	-9.979	-10.165	-10.165	-10.168
W3	-11.35	-11.619	-11.619	-12.366
W4	-50.476	-50.476	-50.476	-51.1
W5	-27.045	-27.045	-27.045	-27.146
W6	-82.693	-82.693	-82.693	-82.701
X	-11.75	-12.798	-12.798	-13.303
Y	-6.559	-6.559	-6.559	-7.226
ZA	-34.659	-34.659	-34.659	-34.675
ZB	-15.874	-15.874	-15.874	-15.917
ZC1	-78.014	-78.014	-78.014	-79.15
ZC2	-2.196	-2.199	-2.199	-4.338
ZC3	-13.43	-13.43	-13.43	-16.149
ZD1	-133.736	-143.087	-143.087	-145.092
ZD2	43.339	38.856	38.856	37.087
ZD3	38.383	38.312	38.312	36.915
ZD4	26.408	26.408	26.408	25.943
ZE1	-93.253	-93.253	-93.253	-93.368
ZE2	10.395	10.395	10.395	9.644
ZE3	-0.369	-0.369	-0.369	-3.164
ZF	-122.822	-122.826	-122.826	-123.229

UN=Unstructured
CSH=Heterscedastic Compound Symmetry
FA0(2)=First order autoregressive
RIS=Random Intercept and Slope
See Section 2.2 for details

Table 36: REML Model Discrimination using the Akaike Information Criterion (AIC) and the Schwarz Bayesian Criterion (SBC) for Cmax in Data Sets A through ZF

Data Set	AIC UN	AIC CSH	AIC FA0(2)	AIC RIS	SBC UN	SBC CSH	SBC FA0(2)	SBC RIS
A	-71.6	-71.6	-71.6	-70.6	-78.3	-78.3	-78.3	-76
B	-280	-280	-280	-279	-289	-289	-289	-286
C1	-19.2	-19.4	-19.4	-19.5	-25.9	-26.2	-26.2	-24.9
C2	-48.2	-48.2	-48.2	-47.6	-54.9	-54.9	-54.9	-52.9
D	-78.4	-78.4	-78.4	-77.5	-85.3	-85.3	-85.3	-83.1
E	-11.5	.	-11.7	-10.8	-16.8	.	-17	-15
F	-38.7	-38.7	-38.7	-38.9	-43.1	-43.1	-43.1	-42.3
G	-53.6	-53.6	-53.6	-53.1	-58.6	-58.6	-58.6	-57.2
H	-8	-8	-8	-7.2	-13.7	-13.7	-13.7	-11.7
I1	-46.1	-46.1	-46.1	-46.3	-52.3	-52.3	-52.3	-51.3
I2	-73.8	-73.8	-73.8	-76	-80.1	-80.1	-80.1	-81
J	-61.3	-61.7	-61.7	-61.1	-67.5	-67.9	-67.9	-66
K1	-124	-125	-125	-124	-132	-132	-132	-129
K2	-8.9	-8.9	-8.9	-8.5	-16.2	-16.2	-16.2	-14.3
K3	-107	-108	-108	-107	-115	-115	-115	-113
L1	-73.9	-73.9	-73.9	-73	-81.3	-81.3	-81.3	-78.9
L2	-117	-118	-118	-117	-125	-126	-126	-123
M	-50.8	-50.8	-50.8	-49.8	-56.5	-56.5	-56.5	-54.4
N1	-13.2	-13.2	-13.2	-12.2	-19.8	-19.8	-19.8	-17.5
N2	-36.1	-36.1	-36.1	-36.1	-42.7	-42.7	-42.7	-41.4
O1	-38.4	-38.4	-38.4	-37.4	-44.6	-44.6	-44.6	-42.4
O2	-52.7	-52.7	-52.7	-52.4	-58.9	-58.9	-58.9	-57.3
P	0.5	0.5	0.5	1.3	-5.5	-5.5	-5.5	-3.6
Q1	10	10	10	10.4	3.3	3.3	3.3	5.1
Q2	-33.9	-33.9	-33.9	-33.4	-40.9	-40.9	-40.9	-39
R	-309	-309	-309	-308	-318	-318	-318	-315

UN=Unstructured
CSH=Heterscedastic Compound Symmetry
FA0(2)=First order autoregressive
RIS=Random Intercept and Slope
See Section 2.2 for details

Table 36: REML Model Discrimination using the Akaike Information Criterion (AIC) and the Schwarz Bayesian Criterion (SBC) for Cmax in Data Sets A through ZF

Data Set	AIC UN	AIC CSH	AIC FA0(2)	AIC RIS	SBC UN	SBC CSH	SBC FA0(2)	SBC RIS
S	-441	-441	-441	-441	-451	-451	-451	-449
T	-408	-408	-408	-408	-418	-418	-418	-415
U	-142	-142	-142	-144	-150	-150	-150	-150
V	-91.9	-92.1	-92.1	-91.1	-98.1	-98.3	-98.3	-96.1
W1	-38.9	-39.1	-39.1	-38.5	-46.1	-46.3	-46.3	-44.3
W2	-36.4	-36.9	-36.9	-36.3	-43.6	-44.1	-44.1	-42
W3	-57.5	-57.5	-57.5	-56.6	-64.7	-64.7	-64.7	-62.4
W4	-50	-50	-50	-50.3	-57.2	-57.2	-57.2	-56.1
W5	-35.3	-35.3	-35.3	-34.7	-42.5	-42.5	-42.5	-40.5
W6	-58.9	-59	-59	-58.2	-66.1	-66.2	-66.2	-64
X	-25.1	-25.1	-25.1	-24.1	-30.8	-30.8	-30.8	-28.6
Y	-50.7	-51.1	-51.1	-50.3	-56.4	-56.7	-56.7	-54.8
ZA	-64.9	-64.9	-64.9	-64.2	-70.9	-70.9	-70.9	-68.9
ZB	-47.5	-47.5	-47.5	-47.9	-53.1	-53.1	-53.1	-52.4
ZC1	-94.8	-94.8	-94.8	-94.4	-103	-103	-103	-101
ZC2	-9.6	-9.9	-9.9	-11.3	-17.3	-17.6	-17.6	-17.5
ZC3	-28.2	-28.2	-28.2	-28.4	-35.9	-36	-36	-34.6
ZD1	-161	-167	-167	-167	-168	-174	-174	-173
ZD2	-0.9	-6.2	-6.2	-5.2	-8.4	-13.7	-13.7	-11.2
ZD3	89.4	89	89	89.6	81.9	81.5	81.5	83.6
ZD4	40.1	40.1	40.1	40.9	32.6	32.6	32.6	34.9
ZE1	-130	-130	-130	-129	-138	-138	-138	-135
ZE2	9.6	9.6	9.6	10.2	2.2	2.2	2.2	4.3
ZE3	1.4	1.4	1.4	0.3	-6	-6	-6	-5.6
ZF	-222	-222	-222	-222	-230	-230	-230	-228

UN=Unstructured
CSH=Heterscedastic Compound Symmetry
FA0(2)=First order autoregressive
RIS=Random Intercept and Slope
See Section 2.2 for details

Table 37: Residual *log*−likelihoods for Cmax in Data Sets A through ZF

Data Set	UN	FA0(2)	CSH	RIS
A	-66.59	-66.604	-66.604	-66.608
B	-274.719	-274.888	-274.888	-275.056
C1	-14.189	-14.45	-14.45	-15.532
C2	-43.179	-43.179	-43.179	-43.575
D	-73.369	-73.369	-73.369	-73.525
E	-6.52	-6.721	.	-6.754
F	-33.719	-33.719	-33.719	-34.86
G	-48.564	-48.564	-48.564	-49.13
H	-3.048	-3.048	-3.048	-3.215
I1	-41.092	-41.092	-41.092	-42.338
I2	-68.819	-68.819	-68.819	-71.993
J	-56.265	-56.67	-56.67	-57.059
K1	-119.339	-119.626	-119.626	-119.63

UN=Unstructured
CSH=Heterscedastic Compound Symmetry
FA0(2)=First order autoregressive
RIS=Random Intercept and Slope
See Section 2.2 for details

Table 37: Residual *log*−likelihoods for Cmax in Data Sets A through ZF

Data Set	UN	FA0(2)	CSH	RIS
K2	-3.924	-3.937	-3.937	-4.502
K3	-102.332	-102.735	-102.735	-102.747
L1	-68.884	-68.925	-68.925	-69
L2	-112.279	-113.224	-113.224	-113.275
M	-45.785	-45.785	-45.785	-45.814
N1	-8.158	-8.158	-8.158	-8.227
N2	-31.123	-31.123	-31.123	-32.067
O1	-33.395	-33.395	-33.395	-33.4
O2	-47.697	-47.697	-47.697	-48.355
P	5.522	5.522	5.522	5.252
Q1	14.954	14.954	14.954	14.416
Q2	-28.89	-28.89	-28.89	-29.424
R	-303.511	-303.639	-303.639	-303.682
S	-436.287	-436.365	-436.365	-437.256
T	-403.451	-403.451	-403.451	-403.541
U	-137.06	-137.163	-137.163	-139.606
V	-86.862	-87.086	-87.086	-87.087
W1	-33.9	-34.111	-34.111	-34.506
W2	-31.397	-31.852	-31.852	-32.271
W3	-52.474	-52.474	-52.474	-52.592
W4	-44.968	-44.971	-44.968	-46.31
W5	-30.301	-30.301	-30.301	-30.745
W6	-53.94	-53.963	-53.963	-54.195
X	-20.091	-20.091	-20.091	-20.096
Y	-45.745	-46.084	-46.084	-46.321
ZA	-59.903	-59.903	-59.903	-60.164
ZB	-42.526	-42.526	-42.526	-43.921
ZC1	-89.817	-89.841	-89.841	-90.431
ZC2	-4.591	-4.865	-4.865	-7.285
ZC3	-23.181	-23.219	-23.219	-24.41
ZD1	-155.679	-161.999	-161.999	-163.259
ZD2	4.079	-1.197	-1.197	-1.197
ZD3	94.408	93.985	93.985	93.581
ZD4	45.098	45.098	45.098	44.912
ZE1	-125.473	-125.473	-125.473	-125.478
ZE2	14.618	14.618	14.618	14.19
ZE3	6.404	6.404	6.404	4.326
ZF	-216.921	-216.921	-216.921	-217.549

UN=Unstructured
CSH=Heterscedastic Compound Symmetry
FA0(2)=First order autoregressive
RIS=Random Intercept and Slope
See Section 2.2 for details

Table 38: Method-of-Moment Analysis for AUC in Data Sets A through ZF

Data	$e^{\hat{\mu}_T-\hat{\mu}_R}$	$\hat{\sigma}^2_{\bar{T}}$	$\hat{\sigma}^2_{\bar{R}}$	$\hat{\sigma}^2_{WT}$	$\hat{\sigma}^2_{WR}$	$\hat{\sigma}^2_I$	$\hat{\sigma}^2_T$	$\hat{\sigma}^2_R$	$\hat{\sigma}^2_D$
A	0.908	0.181	0.203	0.057	0.074	0.061	0.210	0.240	-0.004
B	0.910	1.888	1.934	0.160	0.173	0.150	1.968	2.020	-0.016
C1	1.059	0.028	0.042	0.028	0.048	0.027	0.042	0.066	-0.011
C2	1.069	0.097	0.148	0.028	0.080	0.087	0.111	0.188	0.033
D	0.927	0.295	0.304	0.015	0.028	0.026	0.303	0.318	0.004
E	1.035	0.057	0.057	0.012	0.006	0.010	0.063	0.060	0.001
F	1.007	0.248	0.303	0.048	0.062	0.019	0.272	0.334	-0.036
G	0.820	0.217	0.300	0.030	0.027	0.038	0.231	0.313	0.010
H	1.019	0.086	0.075	0.012	0.017	0.021	0.092	0.084	0.006
I1	0.794	0.124	0.151	0.064	0.032	0.048	0.156	0.167	0.000
I2	0.667	0.262	0.448	0.121	0.087	0.143	0.323	0.492	0.039
J	0.879	0.145	0.178	0.059	0.046	0.061	0.174	0.201	0.009
K1	1.024	0.208	0.236	0.046	0.066	0.045	0.231	0.269	-0.010
K2	0.974	0.041	0.025	0.015	0.018	0.020	0.048	0.034	0.003
K3	1.057	0.061	0.067	0.032	0.031	0.013	0.078	0.082	-0.018
L1	0.920	0.150	0.133	0.061	0.054	0.053	0.181	0.159	-0.004
L2	0.871	0.354	0.360	0.111	0.130	0.076	0.409	0.426	-0.045
M	0.990	0.270	0.250	0.023	0.047	0.037	0.282	0.274	0.002
N1	0.985	0.027	0.018	0.046	0.005	0.020	0.050	0.020	-0.005
N2	0.951	0.110	0.141	0.049	0.044	0.057	0.135	0.163	0.011
O1	0.980	0.069	0.050	0.032	0.026	0.036	0.085	0.063	0.007
O2	1.058	0.199	0.136	0.024	0.075	0.123	0.211	0.174	0.073
P	0.980	0.063	0.066	0.010	0.015	0.007	0.068	0.073	-0.005
Q1	0.900	0.049	0.027	0.014	0.012	0.025	0.056	0.033	0.012
Q2	1.188	0.084	0.069	0.057	0.032	0.053	0.112	0.085	0.009
R	0.918	2.115	2.047	0.295	0.122	0.225	2.263	2.108	0.016
S	1.010	1.684	1.891	0.333	0.204	0.294	1.850	1.993	0.026
T	0.818	1.515	1.322	0.345	0.210	0.391	1.687	1.426	0.113
U	0.989	0.388	0.336	0.141	0.190	0.121	0.459	0.431	-0.045
V	1.020	0.258	0.292	0.057	0.137	0.122	0.287	0.361	0.025
W1	0.996	0.127	0.113	0.035	0.018	0.021	0.145	0.122	-0.005
W2	0.953	0.127	0.116	0.034	0.024	0.026	0.144	0.128	-0.003
W3	0.933	0.206	0.159	0.035	0.020	0.025	0.224	0.168	-0.002
W4	1.094	0.158	0.207	0.044	0.044	0.115	0.180	0.229	0.071
W5	0.984	0.106	0.119	0.033	0.037	0.066	0.122	0.137	0.031
W6	0.995	0.370	0.352	0.065	0.075	0.123	0.403	0.390	0.052
X	1.081	0.130	0.093	0.050	0.037	0.026	0.155	0.112	-0.017
Y	0.933	0.134	0.098	0.021	0.033	0.047	0.145	0.115	0.020
ZA	1.137	0.135	0.172	0.029	0.103	0.074	0.150	0.223	0.008
ZB	1.010	0.063	0.070	0.046	0.034	0.064	0.086	0.087	0.024
ZC1	1.095	0.140	0.185	0.093	0.065	0.096	0.186	0.218	0.017
ZC2	1.056	0.057	0.092	0.029	0.033	0.035	0.072	0.108	0.004
ZC3	1.151	0.102	0.151	0.035	0.024	0.040	0.120	0.163	0.011
ZD1	0.743	1.222	1.524	0.167	0.146	0.066	1.306	1.597	-0.090
ZD2	0.934	0.153	0.179	0.013	0.008	0.007	0.160	0.183	-0.004
ZD3	0.989	0.042	0.055	0.020	0.013	0.017	0.052	0.062	0.000
ZD4	0.991	0.062	0.069	0.019	0.015	0.020	0.072	0.077	0.003
ZE1	0.958	0.447	0.395	0.098	0.067	0.132	0.496	0.429	0.049
ZE2	0.989	0.082	0.055	0.036	0.016	0.030	0.100	0.063	0.004

$\hat{\mu}_t$ =MoM estimate for Formulation Mean (t=Test, Reference)

$\hat{\sigma}^2_{\bar{t}}$ =MoM estimate for Variance of a mean observation (within-subject)

$\hat{\sigma}^2_{Wt}$ =MoM estimate for Within-Subject variance

$\hat{\sigma}^2_I$ =MoM estimate for Variance of $\hat{\mu}_{Ti} - \hat{\mu}_{Ri}$ $(i = 1, \ldots, n)$

$\hat{\sigma}^2_t$ =Derived Total Variance of an observation $(\hat{\sigma}^2_{\bar{t}} + \frac{\hat{\sigma}^2_{Wt}}{2})$

$\hat{\sigma}^2_D$ =Derived estimate for Subject-by-Formulation variance $(\hat{\sigma}^2_I - \frac{\hat{\sigma}^2_{WT}+\hat{\sigma}^2_{WR}}{2})$

Table 38: Method-of-Moment Analysis for AUC in Data Sets A through ZF

Data	$e^{\hat{\mu}_T - \hat{\mu}_R}$	$\hat{\sigma}^2_{\bar{T}}$	$\hat{\sigma}^2_{\bar{R}}$	$\hat{\sigma}^2_{WT}$	$\hat{\sigma}^2_{WR}$	$\hat{\sigma}^2_I$	$\hat{\sigma}^2_T$	$\hat{\sigma}^2_R$	$\hat{\sigma}^2_D$
ZE3	0.926	0.151	0.103	0.021	0.024	0.033	0.161	0.115	0.011
ZF	1.101	0.248	0.305	0.075	0.118	0.098	0.286	0.364	0.002

$\hat{\mu}_t$ =MoM estimate for Formulation Mean (t=Test, Reference)

$\hat{\sigma}^2_{\bar{t}}$ =MoM estimate for Variance of a mean observation (within-subject)

$\hat{\sigma}^2_{Wt}$ =MoM estimate for Within-Subject variance

$\hat{\sigma}^2_I$ =MoM estimate for Variance of $\hat{\mu}_{Ti} - \hat{\mu}_{Ri}$ $(i = 1, \ldots, n)$

$\hat{\sigma}^2_t$ =Derived Total Variance of an observation $(\hat{\sigma}^2_{\bar{t}} + \frac{\hat{\sigma}^2_{Wt}}{2})$

$\hat{\sigma}^2_D$ =Derived estimate for Subject-by-Formulation variance $(\hat{\sigma}^2_I - \frac{\hat{\sigma}^2_{WT} + \hat{\sigma}^2_{WR}}{2})$

Table 39: Method-of-Moment Analysis for Cmax in Data Sets A through ZF

Data	$e^{\hat{\mu}_T-\hat{\mu}_R}$	$\hat{\sigma}^2_{\bar{T}}$	$\hat{\sigma}^2_{\bar{R}}$	$\hat{\sigma}^2_{WT}$	$\hat{\sigma}^2_{WR}$	$\hat{\sigma}^2_I$	$\hat{\sigma}^2_T$	$\hat{\sigma}^2_R$	$\hat{\sigma}^2_D$
A	0.863	0.184	0.193	0.100	0.109	0.103	0.234	0.247	-0.002
B	0.916	0.942	0.908	0.151	0.197	0.156	1.018	1.006	-0.018
C1	1.010	0.057	0.046	0.030	0.057	0.044	0.071	0.075	0.001
C2	1.033	0.093	0.155	0.041	0.087	0.107	0.113	0.198	0.044
D	0.905	0.208	0.236	0.085	0.074	0.143	0.251	0.274	0.063
E	0.904	0.039	0.029	0.070	0.044	0.045	0.074	0.051	-0.012
F	1.027	0.151	0.348	0.121	0.144	0.201	0.212	0.420	0.068
G	0.637	0.196	0.274	0.220	0.107	0.204	0.306	0.328	0.040
H	0.994	0.083	0.103	0.021	0.028	0.037	0.094	0.118	0.012
I1	0.863	0.136	0.189	0.102	0.047	0.086	0.187	0.212	0.012
I2	0.671	0.234	0.463	0.145	0.094	0.196	0.307	0.510	0.076
J	0.969	0.153	0.181	0.138	0.079	0.088	0.222	0.221	-0.021
K1	1.014	0.388	0.425	0.145	0.225	0.148	0.460	0.538	-0.037
K2	0.974	0.073	0.050	0.039	0.030	0.035	0.092	0.065	0.001
K3	1.095	0.305	0.275	0.170	0.125	0.114	0.390	0.338	-0.034
L1	0.936	0.144	0.161	0.079	0.090	0.082	0.184	0.206	-0.003
L2	0.823	0.355	0.393	0.137	0.178	0.104	0.423	0.482	-0.053
M	1.035	0.221	0.193	0.094	0.077	0.141	0.268	0.232	0.056
N1	0.942	0.062	0.034	0.079	0.015	0.061	0.102	0.042	0.014
N2	0.958	0.093	0.131	0.056	0.037	0.078	0.120	0.149	0.031
O1	1.030	0.077	0.079	0.063	0.068	0.085	0.108	0.113	0.019
O2	1.088	0.194	0.144	0.042	0.101	0.150	0.215	0.195	0.078
P	0.846	0.031	0.041	0.026	0.024	0.055	0.044	0.053	0.030
Q1	1.145	0.029	0.051	0.017	0.026	0.042	0.037	0.064	0.020
Q2	1.255	0.065	0.072	0.066	0.038	0.074	0.098	0.092	0.022
R	0.936	1.145	1.111	0.323	0.152	0.215	1.306	1.187	-0.023
S	0.963	0.975	1.102	0.354	0.266	0.298	1.152	1.235	-0.012
T	0.804	0.768	0.680	0.275	0.190	0.346	0.905	0.775	0.114
U	0.977	0.635	0.464	0.108	0.258	0.190	0.689	0.593	0.007
V	0.977	0.267	0.308	0.205	0.299	0.189	0.370	0.458	-0.063
W1	0.961	0.143	0.107	0.067	0.037	0.043	0.176	0.126	-0.009
W2	0.950	0.140	0.106	0.064	0.038	0.039	0.172	0.125	-0.012
W3	0.880	0.186	0.156	0.064	0.048	0.088	0.218	0.180	0.032
W4	1.049	0.134	0.185	0.053	0.056	0.062	0.161	0.213	0.007
W5	1.007	0.101	0.126	0.041	0.050	0.047	0.121	0.151	0.001
W6	0.959	0.164	0.189	0.054	0.074	0.060	0.191	0.226	-0.004
X	1.096	0.113	0.089	0.063	0.029	0.069	0.144	0.103	0.023
Y	0.895	0.116	0.080	0.125	0.145	0.098	0.178	0.152	-0.037
ZA	1.128	0.225	0.231	0.053	0.220	0.134	0.251	0.341	-0.002
ZB	1.116	0.204	0.094	0.081	0.073	0.178	0.244	0.131	0.101
ZC1	1.106	0.108	0.155	0.105	0.114	0.109	0.161	0.212	0.000
ZC2	1.034	0.056	0.093	0.032	0.035	0.033	0.071	0.111	0.000
ZC3	1.102	0.103	0.137	0.037	0.039	0.038	0.121	0.156	0.000
ZD1	0.788	1.160	1.441	0.235	0.221	0.105	1.277	1.551	-0.122
ZD2	0.995	0.088	0.089	0.033	0.036	0.013	0.105	0.107	-0.022
ZD3	0.979	0.012	0.020	0.005	0.013	0.009	0.015	0.027	-0.001
ZD4	0.963	0.042	0.046	0.016	0.013	0.016	0.050	0.052	0.002
ZE1	0.900	0.548	0.537	0.171	0.123	0.196	0.634	0.598	0.049
ZE2	1.005	0.075	0.054	0.030	0.015	0.030	0.090	0.061	0.008

$\hat{\mu}_t$ =MoM estimate for Formulation Mean (t=Test, Reference)
$\hat{\sigma}^2_{\bar{t}}$ =MoM estimate for Variance of a mean observation (within-subject)
$\hat{\sigma}^2_{Wt}$ =MoM estimate for Within-Subject variance
$\hat{\sigma}^2_I$ =MoM estimate for Variance of $\hat{\mu}_{Ti} - \hat{\mu}_{Ri}$ $(i = 1, \ldots, n)$
$\hat{\sigma}^2_t$ =Derived Total Variance of an observation $(\hat{\sigma}^2_{\bar{t}} + \frac{\hat{\sigma}^2_{Wt}}{2})$
$\hat{\sigma}^2_D$ =Derived estimate for Subject-by-Formulation variance $(\hat{\sigma}^2_I - \frac{\hat{\sigma}^2_{WT}+\hat{\sigma}^2_{WR}}{2})$

Table 39: Method-of-Moment Analysis for Cmax in Data Sets A through ZF

Data	$e^{\hat{\mu}_T-\hat{\mu}_R}$	$\hat{\sigma}^2_{\bar{T}}$	$\hat{\sigma}^2_{\bar{R}}$	$\hat{\sigma}^2_{WT}$	$\hat{\sigma}^2_{WR}$	$\hat{\sigma}^2_I$	$\hat{\sigma}^2_T$	$\hat{\sigma}^2_R$	$\hat{\sigma}^2_D$
ZE3	0.928	0.141	0.104	0.020	0.021	0.026	0.152	0.114	0.005
ZF	1.493	0.455	0.355	0.270	0.310	0.340	0.590	0.510	0.050

$\hat{\mu}_t$ =MoM estimate for Formulation Mean (t=Test, Reference)

$\hat{\sigma}^2_{\bar{t}}$ =MoM estimate for Variance of a mean observation (within-subject)

$\hat{\sigma}^2_{Wt}$ =MoM estimate for Within-Subject variance

$\hat{\sigma}^2_I$ =MoM estimate for Variance of $\hat{\mu}_{Ti} - \hat{\mu}_{Ri}$ $(i = 1, \ldots, n)$

$\hat{\sigma}^2_t$ =Derived Total Variance of an observation $(\hat{\sigma}^2_{\bar{t}} + \frac{\hat{\sigma}^2_{Wt}}{2})$

$\hat{\sigma}^2_D$ =Derived estimate for Subject-by-Formulation variance $(\hat{\sigma}^2_I - \frac{\hat{\sigma}^2_{WT}+\hat{\sigma}^2_{WR}}{2})$

Table 40: Unstructued (UN) REML Analysis for AUC in Data Sets A through ZF

Data	$e^{\hat{\mu}_T-\hat{\mu}_R}$	$\hat{\sigma}^2_{BT}$	$\hat{\sigma}^2_{BR}$	COV	$\hat{\sigma}^2_{WT}$	$\hat{\sigma}^2_{WR}$	$\hat{\sigma}^2_{T}$	$\hat{\sigma}^2_{R}$	$\hat{\sigma}^2_{D}$
A	0.914	0.137	0.168	0.158	0.060	0.071	0.197	0.239	-0.010
B	0.906	1.810	1.836	1.838	0.162	0.176	1.972	2.012	-0.029
C1	1.059	0.015	0.017	0.021	0.026	0.049	0.041	0.066	-0.010
C2	1.075	0.084	0.109	0.079	0.027	0.077	0.111	0.187	0.035
D	0.927	0.288	0.289	0.287	0.016	0.029	0.303	0.318	0.003
E	1.036	0.051	0.054	0.052	0.012	0.006	0.063	0.060	0.001
F	1.007	0.224	0.271	0.266	0.049	0.063	0.273	0.334	-0.037
G	0.820	0.204	0.296	0.243	0.029	0.026	0.232	0.323	0.014
H	1.019	0.079	0.068	0.071	0.012	0.016	0.091	0.083	0.005
I1	0.794	0.094	0.135	0.114	0.061	0.031	0.155	0.167	0.002
I2	0.667	0.201	0.404	0.284	0.122	0.088	0.323	0.492	0.038
J	0.879	0.110	0.154	0.130	0.069	0.051	0.179	0.205	0.004
K1	1.024	0.185	0.203	0.199	0.045	0.064	0.230	0.267	-0.011
K2	0.974	0.033	0.017	0.023	0.015	0.017	0.047	0.034	0.003
K3	1.057	0.045	0.052	0.057	0.033	0.030	0.078	0.082	-0.018
L1	0.915	0.118	0.105	0.113	0.061	0.053	0.178	0.158	-0.003
L2	0.864	0.292	0.289	0.314	0.112	0.128	0.404	0.417	-0.047
M	0.990	0.261	0.226	0.242	0.021	0.049	0.282	0.275	0.004
N1	0.985	0.004	0.015	0.012	0.048	0.005	0.052	0.020	-0.006
N2	0.946	0.085	0.113	0.096	0.050	0.045	0.135	0.158	0.006
O1	0.980	0.052	0.037	0.042	0.032	0.027	0.083	0.064	0.004
O2	1.070	0.191	0.101	0.105	0.024	0.072	0.215	0.173	0.082
P	0.982	0.059	0.057	0.060	0.009	0.013	0.069	0.071	-0.004
Q1	0.902	0.043	0.019	0.025	0.014	0.013	0.057	0.032	0.011
Q2	1.198	0.048	0.051	0.047	0.065	0.033	0.113	0.084	0.003
R	0.905	1.962	1.982	1.964	0.285	0.119	2.246	2.102	0.017
S	1.001	1.527	1.743	1.638	0.339	0.211	1.866	1.953	-0.006
T	0.817	1.337	1.217	1.221	0.349	0.211	1.686	1.428	0.112
U	0.976	0.325	0.202	0.294	0.146	0.207	0.471	0.409	-0.061
V	1.024	0.234	0.201	0.210	0.057	0.140	0.292	0.342	0.015
W1	0.996	0.103	0.103	0.107	0.037	0.019	0.139	0.122	-0.007
W2	0.955	0.104	0.103	0.106	0.036	0.026	0.139	0.129	-0.005
W3	0.936	0.176	0.148	0.164	0.038	0.020	0.214	0.169	-0.005
W4	1.086	0.132	0.183	0.124	0.044	0.044	0.176	0.227	0.066
W5	0.983	0.087	0.100	0.079	0.032	0.038	0.119	0.138	0.030
W6	0.987	0.330	0.319	0.288	0.065	0.074	0.395	0.394	0.073
X	1.081	0.104	0.076	0.098	0.049	0.035	0.153	0.110	-0.017
Y	0.941	0.119	0.083	0.093	0.019	0.031	0.139	0.114	0.016
ZA	1.127	0.121	0.111	0.116	0.028	0.124	0.149	0.235	0.001
ZB	0.997	0.041	0.051	0.034	0.045	0.041	0.086	0.091	0.024
ZC1	1.097	0.095	0.152	0.115	0.090	0.064	0.185	0.216	0.017
ZC2	1.057	0.043	0.075	0.057	0.028	0.033	0.071	0.108	0.004
ZC3	1.146	0.086	0.135	0.106	0.033	0.024	0.120	0.159	0.008
ZD1	0.743	1.116	1.436	1.340	0.212	0.177	1.328	1.613	-0.128
ZD2	0.934	0.144	0.174	0.163	0.018	0.010	0.162	0.184	-0.007
ZD3	0.989	0.032	0.048	0.040	0.021	0.014	0.053	0.062	0.000
ZD4	0.989	0.051	0.063	0.056	0.020	0.015	0.071	0.078	0.002
ZE1	0.958	0.399	0.362	0.355	0.096	0.066	0.494	0.428	0.051
ZE2	0.989	0.065	0.047	0.053	0.035	0.016	0.100	0.063	0.005
ZE3	0.926	0.140	0.091	0.110	0.021	0.024	0.161	0.115	0.011

$\hat{\mu}_t$ =REML estimate for Formulation Mean (t=Test, Reference)
$\hat{\sigma}^2_{Bt}$ =REML estimate for Between-subject Variance
$\hat{\sigma}^2_{Wt}$ =REML estimate for Within-Subject variance
COV =REML estimate for covariance of a test and reference obs.
$\hat{\sigma}^2_t$ =Derived Total Variance of an observation ($\hat{\sigma}^2_{Bt}+\hat{\sigma}^2_{Wt}$)
$\hat{\sigma}^2_D$ =Derived Subject-by-Formulation variance ($\hat{\sigma}^2_{BT}+\hat{\sigma}^2_{BR}-2COV$)

Table 40: Unstructued (UN) REML Analysis for AUC in Data Sets A through ZF

Data	$e^{\hat{\mu}_T - \hat{\mu}_R}$	$\hat{\sigma}^2_{BT}$	$\hat{\sigma}^2_{BR}$	COV	$\hat{\sigma}^2_{WT}$	$\hat{\sigma}^2_{WR}$	$\hat{\sigma}^2_{T}$	$\hat{\sigma}^2_{R}$	$\hat{\sigma}^2_{D}$
ZF	1.106	0.205	0.246	0.226	0.076	0.121	0.281	0.367	-0.000

$\hat{\mu}_t$ =REML estimate for Formulation Mean (t=Test, Reference)
$\hat{\sigma}^2_{Bt}$ =REML estimate for Between-subject Variance
$\hat{\sigma}^2_{Wt}$ =REML estimate for Within-Subject variance
COV =REML estimate for covariance of a test and reference obs.
$\hat{\sigma}^2_{t}$ =Derived Total Variance of an observation ($\hat{\sigma}^2_{Bt} + \hat{\sigma}^2_{Wt}$)
$\hat{\sigma}^2_{D}$ =Derived Subject-by-Formulation variance ($\hat{\sigma}^2_{BT} + \hat{\sigma}^2_{BR} - 2COV$)

Table 41: Unstructued (UN) REML Analysis for Cmax in Data Sets A through ZF

Data	$e^{\hat{\mu}_T-\hat{\mu}_R}$	$\hat{\sigma}^2_{BT}$	$\hat{\sigma}^2_{BR}$	$\hat{\rho}\hat{\sigma}_{BT}\hat{\sigma}_{BR}$	$\hat{\sigma}^2_{WT}$	$\hat{\sigma}^2_{WR}$	$\hat{\sigma}^2_{T}$	$\hat{\sigma}^2_{R}$	$\hat{\sigma}^2_{D}$
A	0.862	0.131	0.135	0.136	0.102	0.106	0.232	0.241	-0.005
B	0.916	0.867	0.809	0.847	0.151	0.197	1.018	1.007	-0.019
C1	1.008	0.043	0.010	0.030	0.028	0.071	0.071	0.082	-0.006
C2	1.042	0.073	0.111	0.070	0.039	0.087	0.112	0.198	0.044
D	0.905	0.163	0.198	0.150	0.091	0.077	0.254	0.275	0.060
E	0.913	0.000	0.008	0.010	0.064	0.042	0.064	0.051	-0.012
F	1.027	0.093	0.280	0.149	0.116	0.137	0.209	0.417	0.075
G	0.637	0.092	0.225	0.131	0.210	0.107	0.302	0.332	0.054
H	0.999	0.072	0.087	0.076	0.022	0.028	0.094	0.115	0.008
I1	0.863	0.088	0.166	0.119	0.096	0.046	0.184	0.212	0.015
I2	0.671	0.162	0.416	0.250	0.143	0.093	0.305	0.509	0.078
J	0.969	0.071	0.142	0.122	0.166	0.084	0.237	0.226	-0.030
K1	1.014	0.315	0.314	0.333	0.144	0.218	0.459	0.532	-0.037
K2	0.974	0.052	0.036	0.044	0.038	0.029	0.091	0.065	0.000
K3	1.095	0.217	0.215	0.234	0.174	0.120	0.392	0.335	-0.035
L1	0.928	0.100	0.114	0.110	0.081	0.091	0.181	0.205	-0.006
L2	0.817	0.281	0.300	0.317	0.138	0.175	0.420	0.475	-0.052
M	1.035	0.175	0.155	0.136	0.093	0.076	0.267	0.231	0.058
N1	0.942	0.018	0.026	0.018	0.085	0.015	0.103	0.041	0.009
N2	0.958	0.063	0.109	0.074	0.059	0.038	0.122	0.147	0.024
O1	1.028	0.043	0.047	0.034	0.069	0.066	0.112	0.113	0.022
O2	1.098	0.178	0.097	0.091	0.044	0.098	0.222	0.195	0.093
P	0.848	0.019	0.031	0.009	0.025	0.023	0.044	0.054	0.032
Q1	1.150	0.021	0.036	0.019	0.016	0.025	0.038	0.061	0.018
Q2	1.264	0.024	0.050	0.028	0.074	0.040	0.098	0.090	0.018
R	0.921	0.981	1.027	1.016	0.315	0.151	1.296	1.177	-0.023
S	0.964	0.802	0.963	0.889	0.358	0.269	1.159	1.231	-0.015
T	0.803	0.627	0.585	0.550	0.280	0.193	0.907	0.777	0.112
U	0.978	0.583	0.332	0.454	0.109	0.264	0.692	0.596	0.007
V	0.979	0.174	0.158	0.194	0.195	0.290	0.369	0.448	-0.055
W1	0.961	0.111	0.088	0.104	0.064	0.038	0.176	0.126	-0.008
W2	0.950	0.109	0.087	0.104	0.062	0.038	0.171	0.126	-0.011
W3	0.883	0.151	0.132	0.128	0.063	0.047	0.214	0.179	0.028
W4	1.044	0.103	0.157	0.126	0.051	0.054	0.154	0.212	0.007
W5	1.004	0.077	0.101	0.089	0.041	0.049	0.118	0.150	0.001
W6	0.956	0.129	0.154	0.143	0.057	0.070	0.186	0.224	-0.003
X	1.096	0.078	0.074	0.067	0.065	0.028	0.143	0.102	0.017
Y	0.920	0.055	0.007	0.047	0.129	0.149	0.184	0.156	-0.033
ZA	1.116	0.205	0.127	0.151	0.057	0.228	0.262	0.355	0.030
ZB	1.096	0.177	0.046	0.057	0.080	0.099	0.257	0.145	0.108
ZC1	1.107	0.056	0.097	0.077	0.103	0.113	0.160	0.210	-0.001
ZC2	1.033	0.040	0.074	0.058	0.031	0.036	0.071	0.110	-0.001
ZC3	1.101	0.085	0.117	0.101	0.036	0.038	0.121	0.154	0.000
ZD1	0.788	1.019	1.310	1.247	0.281	0.261	1.300	1.571	-0.165
ZD2	0.995	0.071	0.071	0.082	0.033	0.036	0.105	0.107	-0.022
ZD3	0.979	0.010	0.012	0.012	0.005	0.017	0.015	0.029	-0.002
ZD4	0.963	0.034	0.039	0.036	0.017	0.013	0.050	0.053	0.001
ZE1	0.900	0.465	0.476	0.444	0.167	0.121	0.632	0.597	0.052
ZE2	1.006	0.060	0.047	0.049	0.030	0.014	0.090	0.061	0.009
ZE3	0.928	0.131	0.094	0.110	0.021	0.021	0.152	0.115	0.005

$\hat{\mu}_t$ =REML estimate for Formulation Mean (t=Test, Reference)
$\hat{\sigma}^2_{Bt}$ =REML estimate for Between-subject Variance
$\hat{\sigma}^2_{Wt}$ =REML estimate for Within-Subject variance
COV =REML estimate for covariance of a test and reference obs.
$\hat{\sigma}^2_t$ =Derived Total Variance of an observation $(\hat{\sigma}^2_{Bt}+\hat{\sigma}^2_{Wt})$
$\hat{\sigma}^2_D$ =Derived Subject-by-Formulation variance $(\hat{\sigma}^2_{BT}+\hat{\sigma}^2_{BR}-2COV)$

Table 41: Unstructued (UN) REML Analysis for Cmax in Data Sets A through ZF

Data	$e^{\hat{\mu}_T - \hat{\mu}_R}$	$\hat{\sigma}^2_{BT}$	$\hat{\sigma}^2_{BR}$	$\hat{\rho}\hat{\sigma}_{BT}\hat{\sigma}_{BR}$	$\hat{\sigma}^2_{WT}$	$\hat{\sigma}^2_{WR}$	$\hat{\sigma}^2_{T}$	$\hat{\sigma}^2_{R}$	$\hat{\sigma}^2_{D}$
ZF	1.510	0.311	0.200	0.233	0.269	0.306	0.580	0.506	0.044

$\hat{\mu}_t$ =REML estimate for Formulation Mean (t=Test, Reference)
$\hat{\sigma}^2_{Bt}$ =REML estimate for Between-subject Variance
$\hat{\sigma}^2_{Wt}$ =REML estimate for Within-Subject variance
COV =REML estimate for covariance of a test and reference obs.
$\hat{\sigma}^2_t$ =Derived Total Variance of an observation ($\hat{\sigma}^2_{Bt} + \hat{\sigma}^2_{Wt}$)
$\hat{\sigma}^2_D$ =Derived Subject-by-Formulation variance ($\hat{\sigma}^2_{BT} + \hat{\sigma}^2_{BR} - 2COV$)

Table 42: FA0(2) REML Analysis for AUC in Data Sets A through ZF

Data	$e^{\hat{\mu}_T - \hat{\mu}_R}$	$\hat{\sigma}^2_{BT}$	$\hat{\sigma}^2_{BR}$	$\hat{\sigma}^2_{WT}$	$\hat{\sigma}^2_{WR}$	$\hat{\sigma}^2_D$	$\hat{\sigma}^2_T$	$\hat{\sigma}^2_R$
A	0.913	0.142	0.172	0.056	0.068	0.000	0.198	0.239
B	0.907	1.819	1.848	0.153	0.166	0.000	1.972	2.014
C1	1.060	0.017	0.022	0.024	0.044	0.000	0.041	0.066
C2	1.075	0.084	0.075	0.027	0.077	0.035	0.111	0.152
D	0.927	0.288	0.286	0.016	0.029	0.003	0.303	0.315
E	1.036	0.051	0.054	0.012	0.007	0.000	0.063	0.060
F	1.007	0.235	0.286	0.038	0.049	0.000	0.273	0.335
G	0.820	0.204	0.290	0.029	0.026	0.006	0.232	0.316
H	1.019	0.079	0.063	0.012	0.016	0.004	0.091	0.079
I1	0.794	0.095	0.136	0.060	0.031	0.000	0.155	0.167
I2	0.667	0.201	0.400	0.122	0.088	0.004	0.323	0.487
J	0.879	0.110	0.154	0.069	0.051	0.000	0.179	0.205
K1	1.024	0.188	0.208	0.042	0.059	0.000	0.230	0.267
K2	0.974	0.032	0.017	0.015	0.017	0.000	0.047	0.034
K3	1.057	0.052	0.057	0.027	0.025	0.000	0.078	0.082
L1	0.915	0.119	0.106	0.059	0.051	0.000	0.178	0.158
L2	0.865	0.307	0.307	0.099	0.111	0.000	0.405	0.418
M	0.990	0.261	0.224	0.021	0.049	0.002	0.282	0.272
N1	0.985	0.009	0.015	0.042	0.005	0.000	0.052	0.020
N2	0.946	0.085	0.109	0.050	0.045	0.004	0.135	0.154
O1	0.980	0.052	0.034	0.032	0.027	0.002	0.083	0.062
O2	1.070	0.191	0.058	0.024	0.072	0.043	0.215	0.130
P	0.981	0.060	0.059	0.008	0.012	0.000	0.069	0.071
Q1	0.902	0.043	0.015	0.014	0.013	0.004	0.057	0.028
Q2	1.198	0.048	0.047	0.065	0.033	0.003	0.113	0.080
R	0.905	1.951	1.980	0.295	0.121	0.000	2.246	2.102
S	1.002	1.533	1.746	0.332	0.209	0.000	1.866	1.954
T	0.817	1.337	1.115	0.349	0.211	0.102	1.686	1.326
U	0.978	0.337	0.240	0.132	0.172	0.000	0.469	0.411
V	1.024	0.234	0.188	0.057	0.140	0.013	0.292	0.329
W1	0.996	0.107	0.104	0.033	0.017	0.000	0.140	0.122
W2	0.955	0.106	0.104	0.034	0.025	0.000	0.140	0.129
W3	0.936	0.179	0.149	0.035	0.019	0.000	0.214	0.169
W4	1.086	0.132	0.117	0.044	0.044	0.065	0.176	0.162
W5	0.983	0.087	0.071	0.032	0.038	0.029	0.119	0.109
W6	0.987	0.330	0.252	0.065	0.074	0.068	0.395	0.326
X	1.081	0.111	0.081	0.041	0.030	0.000	0.152	0.111
Y	0.941	0.119	0.073	0.019	0.031	0.010	0.139	0.104
ZA	1.127	0.121	0.110	0.028	0.124	0.001	0.149	0.235
ZB	0.997	0.041	0.028	0.045	0.041	0.023	0.086	0.068
ZC1	1.097	0.095	0.139	0.090	0.064	0.013	0.185	0.203
ZC2	1.057	0.043	0.075	0.028	0.033	0.000	0.071	0.108
ZC3	1.146	0.086	0.131	0.033	0.024	0.004	0.120	0.155
ZD1	0.743	1.175	1.473	0.151	0.142	0.000	1.326	1.616
ZD2	0.934	0.149	0.175	0.013	0.008	0.000	0.162	0.184
ZD3	0.989	0.033	0.048	0.020	0.013	0.000	0.053	0.062
ZD4	0.989	0.051	0.061	0.020	0.015	0.002	0.071	0.076
ZE1	0.958	0.399	0.316	0.096	0.066	0.046	0.494	0.382
ZE2	0.989	0.065	0.044	0.035	0.016	0.003	0.100	0.060
ZE3	0.926	0.140	0.087	0.021	0.024	0.004	0.161	0.111
ZF	1.106	0.206	0.247	0.076	0.120	0.000	0.281	0.367

$\hat{\mu}_t$ =REML estimate for Formulation Mean (t=Test, Reference)
$\hat{\sigma}^2_{Bt}$ =REML estimate for Between-subject Variance
$\hat{\sigma}^2_{Wt}$ =REML estimate for Within-Subject variance
$\hat{\sigma}^2_D$ =REML estimate for Subject-by-Formulation variance
$\hat{\sigma}^2_t$ =Derived Total Variance of an observation ($\hat{\sigma}^2_{Bt} + \hat{\sigma}^2_{Wt}$)

Table 43: FA0(2) REML Analysis for Cmax in Data Sets A through ZF

Data	$e^{\hat{\mu}_T - \hat{\mu}_R}$	$\hat{\sigma}^2_{BT}$	$\hat{\sigma}^2_{BR}$	$\hat{\sigma}^2_{WT}$	$\hat{\sigma}^2_{WR}$	$\hat{\sigma}^2_{D}$	$\hat{\sigma}^2_{T}$	$\hat{\sigma}^2_{R}$
A	0.862	0.132	0.137	0.100	0.104	0.000	0.232	0.242
B	0.916	0.871	0.818	0.147	0.189	0.000	1.018	1.007
C1	1.009	0.044	0.018	0.027	0.063	0.000	0.071	0.082
C2	1.042	0.073	0.067	0.039	0.087	0.044	0.112	0.154
D	0.905	0.163	0.139	0.091	0.077	0.059	0.254	0.216
E	0.912	0.006	0.010	0.059	0.040	0.000	0.065	0.050
F	1.027	0.093	0.238	0.116	0.137	0.041	0.209	0.375
G	0.637	0.092	0.187	0.210	0.107	0.038	0.302	0.294
H	0.999	0.072	0.080	0.022	0.028	0.007	0.094	0.107
I1	0.863	0.088	0.162	0.096	0.046	0.004	0.184	0.208
I2	0.671	0.162	0.386	0.143	0.093	0.031	0.305	0.478
J	0.969	0.094	0.145	0.144	0.080	0.000	0.238	0.225
K1	1.014	0.323	0.331	0.136	0.202	0.000	0.459	0.533
K2	0.974	0.053	0.036	0.038	0.028	0.000	0.091	0.065
K3	1.095	0.233	0.223	0.158	0.112	0.000	0.392	0.335
L1	0.929	0.102	0.117	0.079	0.089	0.000	0.181	0.205
L2	0.818	0.297	0.321	0.124	0.155	0.000	0.421	0.476
M	1.035	0.175	0.106	0.093	0.076	0.049	0.267	0.182
N1	0.942	0.018	0.017	0.085	0.015	0.009	0.103	0.032
N2	0.958	0.063	0.087	0.059	0.038	0.022	0.122	0.124
O1	1.028	0.043	0.027	0.069	0.066	0.020	0.112	0.093
O2	1.098	0.178	0.047	0.044	0.098	0.050	0.222	0.144
P	0.848	0.019	0.004	0.025	0.023	0.027	0.044	0.027
Q1	1.150	0.021	0.018	0.016	0.025	0.018	0.038	0.043
Q2	1.264	0.024	0.033	0.074	0.040	0.017	0.098	0.073
R	0.922	0.995	1.030	0.301	0.147	0.000	1.296	1.177
S	0.964	0.811	0.968	0.348	0.264	0.000	1.159	1.231
T	0.803	0.627	0.482	0.280	0.193	0.102	0.907	0.675
U	0.978	0.584	0.349	0.107	0.248	0.000	0.692	0.597
V	0.978	0.186	0.184	0.183	0.265	0.000	0.369	0.449
W1	0.962	0.115	0.090	0.060	0.036	0.000	0.175	0.126
W2	0.951	0.114	0.090	0.056	0.035	0.000	0.170	0.126
W3	0.883	0.151	0.108	0.063	0.047	0.024	0.214	0.155
W4	1.044	0.102	0.157	0.052	0.055	0.000	0.154	0.212
W5	1.004	0.077	0.101	0.040	0.049	0.000	0.118	0.150
W6	0.956	0.130	0.156	0.056	0.069	0.000	0.186	0.224
X	1.096	0.078	0.058	0.065	0.028	0.016	0.143	0.086
Y	0.917	0.062	0.027	0.122	0.130	0.000	0.183	0.157
ZA	1.116	0.205	0.111	0.057	0.228	0.016	0.262	0.339
ZB	1.096	0.177	0.018	0.080	0.099	0.027	0.257	0.118
ZC1	1.107	0.059	0.099	0.101	0.111	0.000	0.160	0.210
ZC2	1.033	0.042	0.076	0.029	0.034	0.000	0.071	0.110
ZC3	1.101	0.086	0.117	0.035	0.037	0.000	0.121	0.154
ZD1	0.788	1.088	1.362	0.212	0.211	0.000	1.299	1.573
ZD2	0.995	0.078	0.078	0.027	0.029	0.000	0.105	0.107
ZD3	0.979	0.010	0.014	0.005	0.015	0.000	0.015	0.029
ZD4	0.963	0.034	0.038	0.017	0.013	0.001	0.050	0.052
ZE1	0.900	0.465	0.425	0.167	0.121	0.051	0.632	0.546
ZE2	1.006	0.060	0.040	0.030	0.014	0.007	0.090	0.055
ZE3	0.928	0.131	0.092	0.021	0.021	0.002	0.152	0.113
ZF	1.510	0.311	0.175	0.269	0.306	0.025	0.580	0.481

$\hat{\mu}_t$ =REML estimate for Formulation Mean (t=Test, Reference)
$\hat{\sigma}^2_{Bt}$ =REML estimate for Between-subject Variance
$\hat{\sigma}^2_{Wt}$ =REML estimate for Within-Subject variance
$\hat{\sigma}^2_{D}$ =REML estimate for Subject-by-Formulation variance
$\hat{\sigma}^2_{t}$ =Derived Total Variance of an observation ($\hat{\sigma}^2_{Bt} + \hat{\sigma}^2_{Wt}$)

Table 44: CSH REML Analysis for AUC in Data Sets A through ZF

Data	$e^{\hat{\mu}_T - \hat{\mu}_R}$	$\hat{\sigma}^2_{BT}$	$\hat{\sigma}^2_{BR}$	$\hat{\rho}$	$\hat{\sigma}^2_{WT}$	$\hat{\sigma}^2_{WR}$	$\hat{\sigma}^2_D$	$\hat{\sigma}^2_T$	$\hat{\sigma}^2_R$
A	0.913	0.142	0.172	1.000	0.056	0.068	0.001	0.198	0.239
B	0.907	1.819	1.848	1.000	0.153	0.166	0.000	1.972	2.014
C1	1.060	0.017	0.022	1.000	0.024	0.044	0.000	0.041	0.066
C2	1.075	0.084	0.109	0.826	0.027	0.077	0.035	0.111	0.187
D	0.927	0.287	0.289	0.994	0.016	0.029	0.003	0.303	0.318
E	1.036	0.051	0.054	0.993	0.012	0.006	0.001	0.063	0.060
F	1.007	0.235	0.286	1.000	0.038	0.049	0.002	0.273	0.335
G	0.820	0.204	0.296	0.989	0.029	0.026	0.014	0.232	0.323
H	1.019	0.079	0.068	0.968	0.012	0.016	0.005	0.091	0.083
I1	0.794	0.095	0.136	1.000	0.060	0.031	0.004	0.155	0.167
I2	0.667	0.201	0.404	0.995	0.122	0.088	0.038	0.323	0.492
J	0.879	0.110	0.154	1.000	0.069	0.051	0.004	0.179	0.205
K1	1.024	0.188	0.208	1.000	0.042	0.059	0.001	0.230	0.267
K2	0.974	0.032	0.017	0.998	0.015	0.017	0.003	0.047	0.034
K3	1.057	0.052	0.057	1.000	0.027	0.025	0.000	0.078	0.082
L1	0.915	0.119	0.106	1.000	0.059	0.051	0.000	0.178	0.158
L2	0.865	0.307	0.307	1.000	0.099	0.111	0.000	0.405	0.418
M	0.990	0.261	0.226	0.995	0.021	0.049	0.004	0.282	0.275
N1	0.985	0.009	0.015	1.000	0.042	0.005	0.001	0.052	0.020
N2	0.946	0.085	0.113	0.981	0.050	0.045	0.006	0.135	0.158
O1	0.980	0.052	0.037	0.968	0.032	0.027	0.004	0.083	0.064
O2	1.070	0.191	0.101	0.755	0.024	0.072	0.082	0.215	0.173
P	0.981	0.060	0.059	1.000	0.008	0.012	0.000	0.069	0.071
Q1	0.902	0.043	0.019	0.885	0.014	0.013	0.011	0.057	0.032
Q2	1.198	0.048	0.051	0.966	0.065	0.033	0.003	0.113	0.084
R	0.905	1.962	1.982	0.996	0.285	0.119	0.017	2.246	2.102
S	1.002	1.533	1.746	1.000	0.332	0.209	0.007	1.866	1.954
T	0.817	1.337	1.217	0.957	0.349	0.211	0.112	1.686	1.428
U	0.978	0.337	0.240	1.000	0.132	0.172	0.008	0.469	0.411
V	1.024	0.234	0.201	0.967	0.057	0.140	0.015	0.292	0.342
W1	0.996	0.107	0.104	1.000	0.033	0.017	0.000	0.140	0.122
W2	0.955	0.106	0.104	1.000	0.034	0.025	0.000	0.140	0.129
W3	0.936	0.179	0.149	1.000	0.035	0.019	0.001	0.214	0.169
W4	1.086	0.132	0.183	0.801	0.044	0.044	0.066	0.176	0.227
W5	0.983	0.087	0.100	0.843	0.032	0.038	0.030	0.119	0.138
W6	0.987	0.330	0.319	0.888	0.065	0.074	0.073	0.395	0.394
X	1.081	0.111	0.081	1.000	0.041	0.030	0.002	0.152	0.111
Y	0.941	0.119	0.083	0.938	0.019	0.031	0.016	0.139	0.114
ZA	1.127	0.121	0.111	0.997	0.028	0.124	0.001	0.149	0.235
ZB	0.997	0.041	0.051	0.740	0.045	0.041	0.024	0.086	0.091
ZC1	1.097	0.095	0.152	0.955	0.090	0.064	0.017	0.185	0.216
ZC2	1.057	0.043	0.075	1.000	0.028	0.033	0.004	0.071	0.107
ZC3	1.146	0.086	0.135	0.986	0.033	0.024	0.008	0.120	0.159
ZD1	0.743	1.175	1.473	1.000	0.151	0.142	0.017	1.326	1.616
ZD2	0.934	0.149	0.175	1.000	0.013	0.008	0.001	0.162	0.184
ZD3	0.989	0.033	0.048	1.000	0.020	0.013	0.002	0.053	0.062
ZD4	0.989	0.051	0.063	0.987	0.020	0.015	0.002	0.071	0.078
ZE1	0.958	0.399	0.362	0.934	0.096	0.066	0.051	0.494	0.428
ZE2	0.989	0.065	0.047	0.971	0.035	0.016	0.005	0.100	0.063
ZE3	0.926	0.140	0.091	0.976	0.021	0.024	0.011	0.161	0.115

$\hat{\mu}_t$ =REML estimate for Formulation Mean (t=Test, Reference)
$\hat{\sigma}^2_{Bt}$ =REML estimate for Between-subject Variance
$\hat{\sigma}^2_{Wt}$ =REML estimate for Within-Subject variance
$\hat{\rho}$ =REML estimate for correlation of a test and reference obs.
$\hat{\sigma}^2_t$ =Derived Total Variance of an observation ($\hat{\sigma}^2_{Bt} + \hat{\sigma}^2_{Wt}$)
$\hat{\sigma}^2_D$ =Derived Subject-by-Formulation variance ($\hat{\sigma}^2_{BT} + \hat{\sigma}^2_{BR} - 2\hat{\rho}\hat{\sigma}_{BT}\hat{\sigma}_{BR}$)

Table 44: CSH REML Analysis for AUC in Data Sets A through ZF

Data	$e^{\hat{\mu}_T - \hat{\mu}_R}$	$\hat{\sigma}^2_{BT}$	$\hat{\sigma}^2_{BR}$	$\hat{\rho}$	$\hat{\sigma}^2_{WT}$	$\hat{\sigma}^2_{WR}$	$\hat{\sigma}^2_{D}$	$\hat{\sigma}^2_{T}$	$\hat{\sigma}^2_{R}$
ZF	1.106	0.206	0.247	1.000	0.076	0.120	0.002	0.281	0.367

$\hat{\mu}_t$ =REML estimate for Formulation Mean (t=Test, Reference)

$\hat{\sigma}^2_{Bt}$ =REML estimate for Between-subject Variance

$\hat{\sigma}^2_{Wt}$ =REML estimate for Within-Subject variance

$\hat{\rho}$ =REML estimate for correlation of a test and reference obs.

$\hat{\sigma}^2_{t}$ =Derived Total Variance of an observation ($\hat{\sigma}^2_{Bt} + \hat{\sigma}^2_{Wt}$)

$\hat{\sigma}^2_{D}$ =Derived Subject-by-Formulation variance ($\hat{\sigma}^2_{BT} + \hat{\sigma}^2_{BR} - 2\hat{\rho}\hat{\sigma}_{BT}\hat{\sigma}_{BR}$)

Table 45: CSH REML Analysis for Cmax in Data Sets A through ZF

Data	$e^{\hat{\mu}_T - \hat{\mu}_R}$	$\hat{\sigma}^2_{BT}$	$\hat{\sigma}^2_{BR}$	$\hat{\rho}$	$\hat{\sigma}^2_{WT}$	$\hat{\sigma}^2_{WR}$	$\hat{\sigma}^2_D$	$\hat{\sigma}^2_T$	$\hat{\sigma}^2_R$
A	0.862	0.132	0.137	1.000	0.100	0.104	0.000	0.232	0.242
B	0.916	0.871	0.818	1.000	0.147	0.189	0.001	1.018	1.007
C1	1.009	0.044	0.018	1.000	0.027	0.063	0.005	0.071	0.082
C2	1.042	0.073	0.111	0.778	0.039	0.087	0.044	0.112	0.198
D	0.905	0.163	0.198	0.838	0.091	0.077	0.060	0.254	0.275
E	.	0.000	0.008	0.000	0.064	0.042	0.008	0.064	0.050
F	1.027	0.093	0.280	0.923	0.116	0.137	0.075	0.209	0.417
G	0.637	0.092	0.225	0.912	0.210	0.107	0.054	0.302	0.332
H	0.999	0.072	0.087	0.956	0.022	0.028	0.008	0.094	0.115
I1	0.863	0.088	0.166	0.989	0.096	0.046	0.015	0.184	0.212
I2	0.671	0.162	0.416	0.963	0.143	0.093	0.078	0.305	0.509
J	0.969	0.094	0.145	1.000	0.144	0.080	0.006	0.238	0.225
K1	1.014	0.323	0.331	1.000	0.136	0.202	0.000	0.459	0.533
K2	0.974	0.053	0.036	1.000	0.038	0.028	0.002	0.091	0.065
K3	1.095	0.233	0.223	1.000	0.158	0.112	0.000	0.392	0.335
L1	0.929	0.102	0.117	1.000	0.079	0.089	0.000	0.181	0.205
L2	0.818	0.297	0.321	1.000	0.124	0.155	0.000	0.421	0.476
M	1.035	0.175	0.155	0.827	0.093	0.076	0.058	0.267	0.231
N1	0.942	0.018	0.026	0.813	0.085	0.015	0.009	0.103	0.041
N2	0.958	0.063	0.109	0.892	0.059	0.038	0.024	0.122	0.147
O1	1.028	0.043	0.047	0.760	0.069	0.066	0.022	0.112	0.113
O2	1.098	0.178	0.097	0.693	0.044	0.098	0.093	0.222	0.195
P	0.848	0.019	0.031	0.378	0.025	0.023	0.032	0.044	0.054
Q1	1.150	0.021	0.036	0.703	0.016	0.025	0.018	0.038	0.061
Q2	1.264	0.024	0.050	0.812	0.074	0.040	0.018	0.098	0.090
R	0.922	0.995	1.030	1.000	0.301	0.147	0.000	1.296	1.177
S	0.964	0.811	0.967	1.000	0.348	0.264	0.007	1.159	1.231
T	0.803	0.627	0.585	0.908	0.280	0.193	0.112	0.907	0.777
U	0.978	0.584	0.349	1.000	0.107	0.248	0.030	0.692	0.597
V	0.978	0.186	0.184	1.000	0.183	0.265	0.000	0.369	0.449
W1	0.962	0.115	0.090	1.000	0.060	0.036	0.001	0.175	0.126
W2	0.951	0.114	0.090	1.000	0.056	0.035	0.001	0.170	0.126
W3	0.883	0.151	0.132	0.905	0.063	0.047	0.028	0.214	0.179
W4	1.044	0.103	0.157	0.994	0.051	0.054	0.007	0.154	0.212
W5	1.004	0.077	0.101	1.000	0.040	0.049	0.002	0.118	0.150
W6	0.956	0.130	0.156	1.000	0.056	0.069	0.001	0.186	0.224
X	1.096	0.078	0.074	0.885	0.065	0.028	0.017	0.143	0.102
Y	0.917	0.062	0.027	1.000	0.122	0.130	0.007	0.183	0.157
ZA	1.116	0.205	0.127	0.935	0.057	0.228	0.030	0.262	0.355
ZB	1.096	0.177	0.046	0.636	0.080	0.099	0.108	0.257	0.145
ZC1	1.107	0.059	0.099	1.000	0.101	0.111	0.005	0.160	0.210
ZC2	1.033	0.042	0.076	1.000	0.029	0.034	0.005	0.071	0.110
ZC3	1.101	0.086	0.117	1.000	0.035	0.037	0.002	0.121	0.154
ZD1	0.788	1.088	1.362	1.000	0.212	0.211	0.015	1.299	1.573
ZD2	0.995	0.078	0.078	1.000	0.027	0.029	0.000	0.105	0.107
ZD3	0.979	0.010	0.014	1.000	0.005	0.015	0.000	0.015	0.029
ZD4	0.963	0.034	0.039	0.987	0.017	0.013	0.001	0.050	0.053
ZE1	0.900	0.465	0.476	0.945	0.167	0.121	0.052	0.632	0.597
ZE2	1.006	0.060	0.047	0.928	0.030	0.014	0.009	0.090	0.061
ZE3	0.928	0.131	0.094	0.991	0.021	0.021	0.005	0.152	0.115

$\hat{\mu}_t$ =REML estimate for Formulation Mean (t=Test, Reference)

$\hat{\sigma}^2_{Bt}$ =REML estimate for Between-subject Variance

$\hat{\sigma}^2_{Wt}$ =REML estimate for Within-Subject variance

$\hat{\rho}$ =REML estimate for correlation of a test and reference obs.

$\hat{\sigma}^2_t$ =Derived Total Variance of an observation ($\hat{\sigma}^2_{Bt} + \hat{\sigma}^2_{Wt}$)

$\hat{\sigma}^2_D$ =Derived Subject-by-Formulation variance ($\hat{\sigma}^2_{BT} + \hat{\sigma}^2_{BR} - 2\hat{\rho}\hat{\sigma}_{BT}\hat{\sigma}_{BR}$)

Table 45: CSH REML Analysis for Cmax in Data Sets A through ZF

Data	$e^{\hat{\mu}_T - \hat{\mu}_R}$	$\hat{\sigma}^2_{BT}$	$\hat{\sigma}^2_{BR}$	$\hat{\rho}$	$\hat{\sigma}^2_{WT}$	$\hat{\sigma}^2_{WR}$	$\hat{\sigma}^2_D$	$\hat{\sigma}^2_T$	$\hat{\sigma}^2_R$
ZF	1.510	0.311	0.200	0.936	0.269	0.306	0.044	0.580	0.506

$\hat{\mu}_t$ =REML estimate for Formulation Mean (t=Test, Reference)
$\hat{\sigma}^2_{Bt}$ =REML estimate for Between-subject Variance
$\hat{\sigma}^2_{Wt}$ =REML estimate for Within-Subject variance
$\hat{\rho}$ =REML estimate for correlation of a test and reference obs.
$\hat{\sigma}^2_t$ =Derived Total Variance of an observation ($\hat{\sigma}^2_{Bt} + \hat{\sigma}^2_{Wt}$)
$\hat{\sigma}^2_D$ =Derived Subject-by-Formulation variance ($\hat{\sigma}^2_{BT} + \hat{\sigma}^2_{BR} - 2\hat{\rho}\hat{\sigma}_{BT}\hat{\sigma}_{BR}$)

Table 46: Random-Intercept and Random-Slope REML Analysis for AUC in Data Sets A through ZF

Data	$e^{\hat{\mu}_T - \hat{\mu}_R}$	$\hat{\sigma}^2_B$	$\hat{\sigma}^2_D$	$\hat{\sigma}^2_{WT}$	$\hat{\sigma}^2_{WR}$	$\hat{\sigma}^2_T$	$\hat{\sigma}^2_R$
A	0.912	0.155	0.000	0.055	0.070	0.210	0.226
B	0.906	1.833	0.000	0.152	0.166	1.985	1.999
C1	1.060	0.019	0.000	0.024	0.045	0.042	0.064
C2	1.075	0.076	0.030	0.027	0.082	0.103	0.158
D	0.927	0.287	0.003	0.016	0.029	0.302	0.315
E	1.036	0.052	0.001	0.011	0.007	0.064	0.059
F	1.007	0.256	0.000	0.038	0.051	0.294	0.307
G	0.820	0.238	0.014	0.028	0.028	0.266	0.266
H	1.020	0.071	0.005	0.012	0.015	0.083	0.087
I1	0.794	0.118	0.005	0.056	0.033	0.174	0.150
I2	0.667	0.287	0.041	0.109	0.098	0.397	0.385
J	0.879	0.132	0.005	0.064	0.055	0.196	0.186
K1	1.024	0.196	0.000	0.041	0.060	0.237	0.256
K2	0.974	0.023	0.002	0.016	0.016	0.040	0.039
K3	1.057	0.054	0.000	0.026	0.025	0.081	0.080
L1	0.915	0.112	0.000	0.061	0.051	0.173	0.162
L2	0.865	0.307	0.000	0.098	0.111	0.405	0.418
M	0.990	0.247	0.005	0.021	0.047	0.268	0.294
N1	0.985	0.014	0.000	0.042	0.005	0.056	0.019
N2	0.946	0.096	0.005	0.048	0.047	0.144	0.144
O1	0.980	0.042	0.003	0.034	0.026	0.075	0.068
O2	1.070	0.110	0.090	0.024	0.068	0.134	0.177
P	0.981	0.060	0.000	0.008	0.012	0.068	0.071
Q1	0.902	0.025	0.010	0.016	0.012	0.041	0.036
Q2	1.198	0.048	0.004	0.064	0.033	0.112	0.081
R	0.905	1.967	0.017	0.284	0.119	2.251	2.087
S	1.001	1.656	0.000	0.326	0.217	1.982	1.874
T	0.817	1.211	0.106	0.356	0.208	1.567	1.419
U	0.979	0.289	0.000	0.140	0.169	0.429	0.458
V	1.023	0.215	0.021	0.058	0.135	0.272	0.350
W1	0.997	0.105	0.000	0.033	0.017	0.138	0.123
W2	0.955	0.105	0.000	0.034	0.025	0.139	0.130
W3	0.936	0.159	0.000	0.037	0.019	0.196	0.178
W4	1.086	0.124	0.066	0.042	0.046	0.167	0.170
W5	0.983	0.078	0.030	0.032	0.039	0.110	0.117
W6	0.987	0.288	0.073	0.065	0.074	0.354	0.362
X	1.081	0.093	0.000	0.044	0.030	0.136	0.122
Y	0.941	0.096	0.017	0.020	0.029	0.116	0.125
ZA	1.127	0.118	0.004	0.029	0.122	0.146	0.239
ZB	0.997	0.034	0.024	0.044	0.042	0.078	0.076
ZC1	1.097	0.117	0.020	0.083	0.068	0.200	0.185
ZC2	1.057	0.055	0.001	0.026	0.038	0.081	0.093
ZC3	1.145	0.108	0.008	0.031	0.026	0.139	0.134
ZD1	0.743	1.320	0.000	0.155	0.150	1.475	1.471
ZD2	0.934	0.165	0.000	0.013	0.009	0.178	0.173
ZD3	0.989	0.041	0.000	0.019	0.015	0.060	0.056
ZD4	0.990	0.057	0.003	0.019	0.015	0.075	0.072
ZE1	0.958	0.353	0.050	0.097	0.066	0.450	0.419
ZE2	0.989	0.051	0.002	0.038	0.015	0.089	0.066
ZE3	0.926	0.110	0.011	0.023	0.022	0.133	0.133
ZF	1.105	0.221	0.000	0.073	0.125	0.294	0.346

$\hat{\mu}_t$ =REML estimate for Formulation Mean (t=Test, Reference)
$\hat{\sigma}^2_B$ =REML estimate for Between-subject Variance
$\hat{\sigma}^2_D$ =REML estimate for Subject-by-Formulation variance
$\hat{\sigma}^2_{Wt}$ =REML estimate for Within-Subject variance
$\hat{\sigma}^2_t$ =Derived Total Variance of an observation ($\hat{\sigma}^2_B + \hat{\sigma}^2_{Wt}$)

Table 47: Random-Intercept and Random-Slope REML Analysis for Cmax in Data Sets A through ZF

Data	$e^{\hat{\mu}_T - \hat{\mu}_R}$	$\hat{\sigma}^2_B$	$\hat{\sigma}^2_D$	$\hat{\sigma}^2_{WT}$	$\hat{\sigma}^2_{WR}$	$\hat{\sigma}^2_T$	$\hat{\sigma}^2_R$
A	0.862	0.135	0.000	0.099	0.105	0.234	0.240
B	0.916	0.847	0.000	0.149	0.187	0.996	1.034
C1	1.009	0.032	0.003	0.030	0.061	0.063	0.093
C2	1.041	0.067	0.036	0.038	0.094	0.105	0.161
D	0.905	0.151	0.060	0.089	0.079	0.240	0.230
E	0.912	0.008	0.000	0.058	0.042	0.066	0.050
F	1.027	0.139	0.053	0.104	0.166	0.243	0.305
G	0.637	0.138	0.071	0.191	0.114	0.329	0.252
H	0.999	0.075	0.007	0.021	0.029	0.096	0.104
I1	0.863	0.126	0.023	0.087	0.049	0.213	0.175
I2	0.671	0.256	0.085	0.127	0.103	0.384	0.359
J	0.969	0.123	0.000	0.141	0.086	0.264	0.209
K1	1.014	0.327	0.000	0.136	0.202	0.462	0.529
K2	0.974	0.043	0.000	0.041	0.027	0.084	0.070
K3	1.095	0.227	0.000	0.159	0.112	0.387	0.339
L1	0.929	0.109	0.000	0.077	0.091	0.186	0.200
L2	0.818	0.307	0.000	0.123	0.156	0.430	0.464
M	1.035	0.136	0.057	0.094	0.075	0.230	0.211
N1	0.942	0.019	0.013	0.081	0.015	0.100	0.034
N2	0.958	0.076	0.027	0.054	0.040	0.130	0.116
O1	1.028	0.034	0.022	0.068	0.067	0.102	0.101
O2	1.098	0.094	0.101	0.045	0.092	0.140	0.186
P	0.848	0.009	0.032	0.024	0.024	0.033	0.033
Q1	1.150	0.019	0.016	0.016	0.028	0.034	0.047
Q2	1.263	0.029	0.023	0.066	0.043	0.096	0.072
R	0.922	1.018	0.000	0.300	0.148	1.318	1.167
S	0.965	0.895	0.000	0.339	0.276	1.234	1.170
T	0.803	0.548	0.109	0.283	0.191	0.831	0.738
U	0.979	0.484	0.030	0.116	0.236	0.600	0.720
V	0.978	0.185	0.000	0.183	0.265	0.368	0.450
W1	0.963	0.098	0.000	0.063	0.035	0.161	0.133
W2	0.952	0.098	0.000	0.059	0.035	0.157	0.133
W3	0.884	0.127	0.027	0.064	0.046	0.191	0.173
W4	1.043	0.124	0.006	0.048	0.059	0.172	0.184
W5	1.004	0.087	0.000	0.039	0.052	0.126	0.139
W6	0.955	0.141	0.000	0.054	0.071	0.196	0.212
X	1.096	0.067	0.017	0.066	0.028	0.133	0.095
Y	0.917	0.041	0.000	0.131	0.127	0.172	0.168
ZA	1.116	0.161	0.050	0.057	0.213	0.218	0.374
ZB	1.096	0.054	0.113	0.090	0.088	0.144	0.142
ZC1	1.107	0.075	0.000	0.096	0.121	0.171	0.196
ZC2	1.033	0.055	0.000	0.027	0.040	0.082	0.094
ZC3	1.101	0.100	0.000	0.034	0.040	0.133	0.140
ZD1	0.788	1.217	0.000	0.212	0.221	1.429	1.438
ZD2	0.995	0.078	0.000	0.027	0.029	0.105	0.107
ZD3	0.979	0.011	0.000	0.005	0.016	0.016	0.026
ZD4	0.963	0.036	0.001	0.016	0.014	0.052	0.050
ZE1	0.900	0.445	0.053	0.166	0.121	0.611	0.566
ZE2	1.006	0.048	0.007	0.032	0.014	0.079	0.062
ZE3	0.928	0.109	0.005	0.022	0.020	0.131	0.129
ZF	1.511	0.233	0.044	0.287	0.288	0.521	0.521

$\hat{\mu}_t$ =REML estimate for Formulation Mean (t=Test, Reference)
$\hat{\sigma}^2_B$ =REML estimate for Between-subject Variance
$\hat{\sigma}^2_D$ =REML estimate for Subject-by-Formulation variance
$\hat{\sigma}^2_{Wt}$ =REML estimate for Within-Subject variance
$\hat{\sigma}^2_t$ =Derived Total Variance of an observation ($\hat{\sigma}^2_B + \hat{\sigma}^2_{Wt}$)

Table 48: Average Bioequivalence Point Estimates for AUC in Data Sets A through ZF

Data	MoM	UN	CSH	FA0(2)	RIS
A	0.908	0.914	0.913	0.913	0.912
B	0.910	0.906	0.907	0.907	0.906
C1	1.059	1.059	1.060	1.060	1.060
C2	1.069	1.075	1.075	1.075	1.075
D	0.927	0.927	0.927	0.927	0.927
E	1.035	1.036	1.036	1.036	1.036
F	1.007	1.007	1.007	1.007	1.007
G	0.820	0.820	0.820	0.820	0.820
H	1.019	1.019	1.019	1.019	1.020
I1	0.794	0.794	0.794	0.794	0.794
I2	0.667	0.667	0.667	0.667	0.667
J	0.879	0.879	0.879	0.879	0.879
K1	1.024	1.024	1.024	1.024	1.024
K2	0.974	0.974	0.974	0.974	0.974
K3	1.057	1.057	1.057	1.057	1.057
L1	0.920	0.915	0.915	0.915	0.915
L2	0.871	0.864	0.865	0.865	0.865
M	0.990	0.990	0.990	0.990	0.990
N1	0.985	0.985	0.985	0.985	0.985
N2	0.951	0.946	0.946	0.946	0.946
O1	0.980	0.980	0.980	0.980	0.980
O2	1.058	1.070	1.070	1.070	1.070
P	0.980	0.982	0.981	0.981	0.981
Q1	0.900	0.902	0.902	0.902	0.902
Q2	1.188	1.198	1.198	1.198	1.198
R	0.918	0.905	0.905	0.905	0.905
S	1.010	1.001	1.002	1.002	1.001
T	0.818	0.817	0.817	0.817	0.817
U	0.989	0.976	0.978	0.978	0.979
V	1.020	1.024	1.024	1.024	1.023
W1	0.996	0.996	0.996	0.996	0.997
W2	0.953	0.955	0.955	0.955	0.955
W3	0.933	0.936	0.936	0.936	0.936
W4	1.094	1.086	1.086	1.086	1.086
W5	0.984	0.983	0.983	0.983	0.983
W6	0.995	0.987	0.987	0.987	0.987
X	1.081	1.081	1.081	1.081	1.081
Y	0.933	0.941	0.941	0.941	0.941
ZA	1.137	1.127	1.127	1.127	1.127
ZB	1.010	0.997	0.997	0.997	0.997
ZC1	1.095	1.097	1.097	1.097	1.097
ZC2	1.056	1.057	1.057	1.057	1.057
ZC3	1.151	1.146	1.146	1.146	1.145
ZD1	0.743	0.743	0.743	0.743	0.743
ZD2	0.934	0.934	0.934	0.934	0.934
ZD3	0.989	0.989	0.989	0.989	0.989
ZD4	0.991	0.989	0.989	0.989	0.990
ZE1	0.958	0.958	0.958	0.958	0.958
ZE2	0.989	0.989	0.989	0.989	0.989
ZE3	0.926	0.926	0.926	0.926	0.926
ZF	1.101	1.106	1.106	1.106	1.105

MoM: Method-of-Moments Estimate
UN: Unstructured REML Estimate
CSH: Heteroscedastic REML Estimate
FA0(2): First-Order REML Estimate
RIS: Random-Intercept and Slope REML Estimate

Table 49: Average Bioequivalence Lower 90% Bound for AUC in Data Sets A through ZF

Data	MoM	UN	CSH	FA0(2)	RIS
A	0.843	0.848	0.845	0.845	0.844
B	0.845	0.842	0.839	0.839	0.838
C1	1.003	1.006	1.002	1.002	1.001
C2	0.973	0.980	0.980	0.980	0.981
D	0.884	0.884	0.884	0.884	0.884
E	0.992	0.994	0.994	0.996	0.993
F	0.936	0.938	0.907	0.907	0.909
G	0.750	0.751	0.751	0.751	0.751
H	0.962	0.967	0.967	0.967	0.966
I1	0.735	0.735	0.736	0.736	0.734
I2	0.584	0.585	0.585	0.585	0.584
J	0.805	0.806	0.807	0.807	0.805
K1	0.966	0.966	0.962	0.962	0.962
K2	0.937	0.937	0.937	0.938	0.937
K3	1.021	1.023	1.011	1.011	1.011
L1	0.862	0.859	0.858	0.858	0.858
L2	0.808	0.801	0.792	0.792	0.792
M	0.916	0.918	0.918	0.918	0.918
N1	0.939	0.940	0.938	0.938	0.938
N2	0.883	0.879	0.879	0.879	0.878
O1	0.919	0.919	0.919	0.919	0.920
O2	0.935	0.943	0.943	0.943	0.941
P	0.949	0.951	0.947	0.947	0.947
Q1	0.855	0.858	0.858	0.858	0.859
Q2	1.108	1.118	1.118	1.118	1.118
R	0.836	0.825	0.825	0.827	0.825
S	0.922	0.915	0.914	0.914	0.915
T	0.735	0.735	0.735	0.735	0.735
U	0.893	0.890	0.879	0.879	0.882
V	0.904	0.912	0.912	0.912	0.910
W1	0.955	0.956	0.952	0.952	0.953
W2	0.911	0.912	0.910	0.910	0.910
W3	0.892	0.894	0.892	0.892	0.893
W4	0.994	0.987	0.987	0.987	0.986
W5	0.915	0.914	0.914	0.914	0.913
W6	0.901	0.885	0.885	0.885	0.885
X	1.018	1.017	1.005	1.005	1.007
Y	0.862	0.870	0.870	0.870	0.870
ZA	1.017	1.019	1.019	1.019	1.018
ZB	0.909	0.900	0.900	0.900	0.900
ZC1	1.013	1.014	1.014	1.014	1.013
ZC2	1.007	1.008	1.009	1.009	1.009
ZC3	1.095	1.090	1.090	1.090	1.089
ZD1	0.692	0.692	0.666	0.666	0.668
ZD2	0.914	0.914	0.907	0.907	0.908
ZD3	0.954	0.954	0.953	0.953	0.954
ZD4	0.953	0.951	0.951	0.951	0.952
ZE1	0.866	0.866	0.866	0.866	0.866
ZE2	0.942	0.942	0.942	0.942	0.944
ZE3	0.880	0.880	0.880	0.880	0.880
ZF	1.024	1.029	1.029	1.029	1.029

MoM: Method-of-Moments Estimate
UN: Unstructured REML Estimate
CSH: Heteroscedastic REML Estimate
FA0(2): First-Order REML Estimate
RIS: Random-Intercept and Slope REML Estimate

Table 50: Average Bioequivalence Upper 90% Bound for AUC in Data Sets A through ZF

Data	MoM	UN	CSH	FA0(2)	RIS
A	0.978	0.984	0.987	0.987	0.985
B	0.981	0.976	0.980	0.980	0.980
C1	1.117	1.116	1.122	1.122	1.122
C2	1.174	1.180	1.180	1.180	1.178
D	0.972	0.973	0.973	0.973	0.973
E	1.080	1.081	1.081	1.078	1.081
F	1.083	1.080	1.118	1.118	1.116
G	0.897	0.896	0.896	0.896	0.896
H	1.079	1.075	1.075	1.075	1.075
I1	0.857	0.857	0.856	0.856	0.858
I2	0.762	0.761	0.761	0.761	0.762
J	0.961	0.960	0.958	0.958	0.960
K1	1.086	1.086	1.090	1.090	1.090
K2	1.011	1.012	1.012	1.011	1.012
K3	1.094	1.092	1.105	1.105	1.105
L1	0.981	0.976	0.976	0.976	0.976
L2	0.939	0.931	0.945	0.945	0.945
M	1.071	1.068	1.068	1.068	1.069
N1	1.034	1.032	1.035	1.035	1.035
N2	1.025	1.019	1.019	1.019	1.018
O1	1.044	1.045	1.045	1.045	1.044
O2	1.198	1.213	1.213	1.213	1.217
P	1.011	1.014	1.017	1.017	1.017
Q1	0.948	0.949	0.949	0.949	0.948
Q2	1.274	1.284	1.284	1.284	1.284
R	1.008	0.993	0.993	0.990	0.993
S	1.108	1.096	1.097	1.097	1.096
T	0.910	0.909	0.909	0.909	0.908
U	1.095	1.070	1.088	1.088	1.088
V	1.151	1.151	1.151	1.151	1.151
W1	1.038	1.038	1.042	1.042	1.043
W2	0.997	1.001	1.003	1.003	1.003
W3	0.976	0.979	0.982	0.982	0.982
W4	1.204	1.195	1.195	1.195	1.195
W5	1.058	1.058	1.058	1.058	1.058
W6	1.098	1.101	1.101	1.101	1.101
X	1.148	1.149	1.163	1.163	1.161
Y	1.010	1.017	1.017	1.017	1.017
ZA	1.272	1.248	1.248	1.248	1.249
ZB	1.122	1.104	1.104	1.104	1.104
ZC1	1.184	1.187	1.187	1.187	1.187
ZC2	1.107	1.109	1.108	1.108	1.107
ZC3	1.210	1.205	1.205	1.205	1.204
ZD1	0.797	0.797	0.828	0.828	0.825
ZD2	0.955	0.955	0.962	0.962	0.961
ZD3	1.025	1.024	1.025	1.025	1.025
ZD4	1.030	1.028	1.028	1.028	1.029
ZE1	1.059	1.059	1.059	1.059	1.059
ZE2	1.038	1.038	1.038	1.038	1.036
ZE3	0.974	0.974	0.974	0.974	0.974
ZF	1.183	1.188	1.188	1.188	1.187

MoM: Method-of-Moments Estimate
UN: Unstructured REML Estimate
CSH: Heteroscedastic REML Estimate
FA0(2): First-Order REML Estimate
RIS: Random-Intercept and Slope REML Estimate

Table 51: Average Bioequivalence Point Estimates for Cmax in Data Sets A through ZF

Data	MoM	UN	CSH	FA0(2)	RIS
A	0.863	0.862	0.862	0.862	0.862
B	0.916	0.916	0.916	0.916	0.916
C1	1.010	1.008	1.009	1.009	1.009
C2	1.033	1.042	1.042	1.042	1.041
D	0.905	0.905	0.905	0.905	0.905
E	0.904	0.913	.	0.912	0.912
F	1.027	1.027	1.027	1.027	1.027
G	0.637	0.637	0.637	0.637	0.637
H	0.994	0.999	0.999	0.999	0.999
I1	0.863	0.863	0.863	0.863	0.863
I2	0.671	0.671	0.671	0.671	0.671
J	0.969	0.969	0.969	0.969	0.969
K1	1.014	1.014	1.014	1.014	1.014
K2	0.974	0.974	0.974	0.974	0.974
K3	1.095	1.095	1.095	1.095	1.095
L1	0.936	0.928	0.929	0.929	0.929
L2	0.823	0.817	0.818	0.818	0.818
M	1.035	1.035	1.035	1.035	1.035
N1	0.942	0.942	0.942	0.942	0.942
N2	0.958	0.958	0.958	0.958	0.958
O1	1.030	1.028	1.028	1.028	1.028
O2	1.088	1.098	1.098	1.098	1.098
P	0.846	0.848	0.848	0.848	0.848
Q1	1.145	1.150	1.150	1.150	1.150
Q2	1.255	1.264	1.264	1.264	1.263
R	0.936	0.921	0.922	0.922	0.922
S	0.963	0.964	0.964	0.964	0.965
T	0.804	0.803	0.803	0.803	0.803
U	0.977	0.978	0.978	0.978	0.979
V	0.977	0.979	0.978	0.978	0.978
W1	0.961	0.961	0.962	0.962	0.963
W2	0.950	0.950	0.951	0.951	0.952
W3	0.880	0.883	0.883	0.883	0.884
W4	1.049	1.044	1.044	1.044	1.043
W5	1.007	1.004	1.004	1.004	1.004
W6	0.959	0.956	0.956	0.956	0.955
X	1.096	1.096	1.096	1.096	1.096
Y	0.895	0.920	0.917	0.917	0.917
ZA	1.128	1.116	1.116	1.116	1.116
ZB	1.116	1.096	1.096	1.096	1.096
ZC1	1.106	1.107	1.107	1.107	1.107
ZC2	1.034	1.033	1.033	1.033	1.033
ZC3	1.102	1.101	1.101	1.101	1.101
ZD1	0.788	0.788	0.788	0.788	0.788
ZD2	0.995	0.995	0.995	0.995	0.995
ZD3	0.979	0.979	0.979	0.979	0.979
ZD4	0.963	0.963	0.963	0.963	0.963
ZE1	0.900	0.900	0.900	0.900	0.900
ZE2	1.005	1.006	1.006	1.006	1.006
ZE3	0.928	0.928	0.928	0.928	0.928
ZF	1.493	1.510	1.510	1.510	1.511

MoM: Method-of-Moments Estimate
UN: Unstructured REML Estimate
CSH: Heteroscedastic REML Estimate
FA0(2): First-Order REML Estimate
RIS: Random-Intercept and Slope REML Estimate

Table 52: Average Bioequivalence Lower 90% Bound for Cmax in Data Sets A through ZF

Data	MoM	UN	CSH	FA0(2)	RIS
A	0.783	0.781	0.781	0.781	0.781
B	0.849	0.849	0.846	0.846	0.847
C1	0.940	0.944	0.942	0.942	0.942
C2	0.931	0.940	0.940	0.940	0.942
D	0.808	0.808	0.808	0.808	0.808
E	0.828	0.839	.	0.832	0.832
F	0.812	0.817	0.817	0.817	0.824
G	0.519	0.521	0.521	0.521	0.519
H	0.922	0.932	0.932	0.932	0.932
I1	0.779	0.779	0.779	0.779	0.777
I2	0.575	0.575	0.575	0.575	0.574
J	0.868	0.870	0.862	0.862	0.864
K1	0.911	0.911	0.905	0.905	0.905
K2	0.925	0.925	0.925	0.925	0.926
K3	0.996	0.997	0.989	0.989	0.989
L1	0.865	0.859	0.858	0.858	0.858
L2	0.752	0.747	0.739	0.739	0.739
M	0.897	0.896	0.896	0.896	0.896
N1	0.870	0.871	0.871	0.871	0.870
N2	0.879	0.879	0.879	0.879	0.878
O1	0.928	0.926	0.926	0.926	0.926
O2	0.947	0.954	0.954	0.954	0.952
P	0.776	0.779	0.779	0.779	0.779
Q1	1.074	1.079	1.079	1.079	1.081
Q2	1.155	1.165	1.165	1.165	1.162
R	0.855	0.842	0.841	0.841	0.841
S	0.877	0.878	0.876	0.876	0.878
T	0.727	0.726	0.726	0.726	0.727
U	0.867	0.869	0.867	0.867	0.866
V	0.837	0.843	0.833	0.833	0.833
W1	0.906	0.905	0.902	0.902	0.904
W2	0.898	0.898	0.894	0.894	0.895
W3	0.810	0.813	0.813	0.813	0.814
W4	0.978	0.973	0.973	0.975	0.972
W5	0.947	0.944	0.945	0.945	0.944
W6	0.895	0.890	0.890	0.890	0.890
X	0.994	0.994	0.994	0.994	0.994
Y	0.783	0.810	0.799	0.799	0.801
ZA	0.958	0.959	0.959	0.959	0.954
ZB	0.932	0.920	0.920	0.920	0.917
ZC1	1.018	1.018	1.018	1.018	1.019
ZC2	0.987	0.987	0.985	0.985	0.986
ZC3	1.049	1.048	1.048	1.048	1.049
ZD1	0.721	0.721	0.694	0.694	0.695
ZD2	0.965	0.965	0.951	0.951	0.951
ZD3	0.954	0.954	0.952	0.952	0.952
ZD4	0.930	0.930	0.930	0.930	0.930
ZE1	0.796	0.796	0.796	0.796	0.796
ZE2	0.958	0.958	0.958	0.958	0.959
ZE3	0.887	0.887	0.887	0.887	0.888
ZF	1.307	1.323	1.323	1.323	1.324

MoM: Method-of-Moments Estimate
UN: Unstructured REML Estimate
CSH: Heteroscedastic REML Estimate
FA0(2): First-Order REML Estimate
RIS: Random-Intercept and Slope REML Estimate

Table 53: Average Bioequivalence Upper 90% Bound for Cmax in Data Sets A through ZF

Data	MoM	UN	CSH	FA0(2)	RIS
A	0.951	0.951	0.951	0.951	0.951
B	0.989	0.989	0.991	0.991	0.991
C1	1.085	1.076	1.081	1.081	1.081
C2	1.147	1.155	1.155	1.155	1.152
D	1.013	1.013	1.013	1.013	1.014
E	0.988	0.994	.	1.000	0.999
F	1.298	1.290	1.290	1.290	1.279
G	0.781	0.778	0.778	0.778	0.781
H	1.071	1.071	1.071	1.071	1.071
I1	0.957	0.956	0.956	0.956	0.959
I2	0.784	0.783	0.783	0.783	0.785
J	1.081	1.078	1.088	1.088	1.086
K1	1.128	1.128	1.136	1.136	1.136
K2	1.025	1.026	1.025	1.025	1.025
K3	1.203	1.203	1.212	1.212	1.212
L1	1.013	1.004	1.005	1.005	1.005
L2	0.900	0.893	0.905	0.905	0.905
M	1.194	1.196	1.196	1.196	1.196
N1	1.020	1.018	1.018	1.018	1.019
N2	1.044	1.044	1.044	1.044	1.045
O1	1.144	1.140	1.140	1.140	1.140
O2	1.250	1.265	1.265	1.265	1.268
P	0.923	0.922	0.922	0.922	0.922
Q1	1.221	1.225	1.225	1.225	1.224
Q2	1.364	1.372	1.372	1.372	1.373
R	1.025	1.008	1.011	1.011	1.011
S	1.057	1.059	1.060	1.060	1.060
T	0.889	0.888	0.888	0.888	0.887
U	1.101	1.102	1.104	1.104	1.106
V	1.139	1.137	1.147	1.147	1.147
W1	1.018	1.021	1.025	1.025	1.025
W2	1.004	1.006	1.012	1.012	1.013
W3	0.957	0.960	0.960	0.960	0.960
W4	1.126	1.120	1.120	1.119	1.120
W5	1.070	1.068	1.067	1.067	1.068
W6	1.028	1.026	1.026	1.026	1.026
X	1.208	1.208	1.208	1.208	1.207
Y	1.023	1.044	1.052	1.052	1.050
ZA	1.327	1.299	1.299	1.299	1.305
ZB	1.336	1.307	1.307	1.307	1.308
ZC1	1.202	1.204	1.205	1.205	1.204
ZC2	1.082	1.082	1.084	1.084	1.082
ZC3	1.157	1.156	1.157	1.157	1.156
ZD1	0.862	0.862	0.896	0.896	0.894
ZD2	1.026	1.026	1.041	1.041	1.041
ZD3	1.004	1.004	1.006	1.006	1.006
ZD4	0.997	0.997	0.997	0.997	0.997
ZE1	1.017	1.017	1.017	1.017	1.017
ZE2	1.056	1.056	1.056	1.056	1.055
ZE3	0.970	0.970	0.970	0.970	0.970
ZF	1.707	1.723	1.723	1.723	1.725

MoM: Method-of-Moments Estimate
UN: Unstructured REML Estimate
CSH: Heteroscedastic REML Estimate
FA0(2): First-Order REML Estimate
RIS: Random-Intercept and Slope REML Estimate

Table 54: Hyslop et al (2000) Upper 90% Bound for Linearised Population and Individual Bioequivalence FDA Metric of AUC in Data Sets A through ZF

Data	$\hat{\sigma}_R$	$\hat{\nu}_{PBE}$	$\hat{\nu}_{C.PBE}$	$\hat{\sigma}_{WR}$	$\hat{\nu}_{IBE}$	$\hat{\nu}_{C.IBE}$
A	0.490	-0.211	0.056	0.271	-0.084	-0.048
B	1.421	-2.209	0.673	0.415	-0.281	-0.034
C1	0.257	-0.087	-0.063	0.219	-0.080	-0.094
C2	0.433	-0.241	-0.053	0.282	-0.088	-0.037
D	0.564	-0.233	0.138	0.169	-0.032	-0.078
E	0.245	-0.013	0.000	0.080	0.008	-0.079
F	0.578	-0.104	0.277	0.249	-0.082	-0.086
G	0.559	-0.135	0.200	0.163	0.066	0.021
H	0.290	-0.015	0.035	0.132	-0.007	-0.073
I1	0.409	-0.067	0.094	0.179	0.081	0.049
I2	0.701	-0.364	0.217	0.295	0.232	0.317
J	0.449	-0.142	0.063	0.215	0.014	0.007
K1	0.519	-0.271	0.037	0.256	-0.108	-0.086
K2	0.185	-0.011	-0.029	0.133	-0.018	-0.083
K3	0.287	-0.076	-0.028	0.175	-0.051	-0.094
L1	0.399	-0.112	0.052	0.232	-0.050	-0.046
L2	0.652	-0.384	0.158	0.361	-0.203	-0.062
M	0.523	-0.085	0.229	0.216	-0.056	-0.074
N1	0.142	0.026	-0.013	0.073	0.046	-0.042
N2	0.403	-0.152	0.002	0.210	-0.016	-0.029
O1	0.252	-0.009	0.016	0.161	-0.001	-0.049
O2	0.417	-0.042	0.148	0.274	-0.010	0.040
P	0.270	-0.051	-0.019	0.121	-0.023	-0.098
Q1	0.181	0.020	0.001	0.111	0.022	-0.052
Q2	0.291	-0.004	0.053	0.178	0.047	0.014
R	1.452	-2.065	0.954	0.350	0.042	0.203
S	1.412	-2.451	0.422	0.452	-0.158	0.175
T	1.194	-1.328	0.756	0.458	-0.016	0.340
U	0.656	-0.363	0.202	0.436	-0.330	-0.078
V	0.601	-0.341	0.094	0.370	-0.175	-0.028
W1	0.349	-0.064	0.041	0.133	-0.005	-0.068
W2	0.358	-0.079	0.036	0.155	-0.016	-0.068
W3	0.410	-0.048	0.129	0.141	0.000	-0.060
W4	0.479	-0.238	0.010	0.211	0.062	0.059
W5	0.371	-0.133	-0.008	0.192	-0.002	-0.026
W6	0.624	-0.289	0.199	0.275	-0.025	0.029
X	0.334	0.023	0.119	0.193	-0.022	-0.054
Y	0.339	0.015	0.114	0.183	0.004	-0.030
ZA	0.473	-0.221	0.019	0.322	-0.136	-0.052
ZB	0.295	-0.046	0.007	0.184	0.042	0.012
ZC1	0.467	-0.236	0.003	0.255	-0.014	0.019
ZC2	0.329	-0.143	-0.062	0.181	-0.036	-0.071
ZC3	0.403	-0.176	-0.022	0.154	0.021	-0.028
ZD1	1.264	-1.587	0.535	0.382	-0.158	0.019
ZD2	0.428	-0.171	0.011	0.091	-0.004	-0.087
ZD3	0.248	-0.066	-0.050	0.115	-0.006	-0.080
ZD4	0.277	-0.072	-0.034	0.121	-0.008	-0.079
ZE1	0.655	-0.269	0.277	0.260	0.027	0.068
ZE2	0.250	-0.007	0.018	0.127	0.011	-0.055

$\hat{\sigma}_R$: MoM Estimate for Total SD, Reference Formula
$\hat{\nu}_{PBE}$: Upper 90% Bound for Linearised PBE Metric
$\hat{\nu}_{C.PBE}$: Upper 90% Bound for Linearised Constant-Scaled PBE Metric
$\hat{\sigma}_{WR}$: MoM Estimate for Within-Subject SD, Reference Formula
$\hat{\nu}_{IBE}$: Upper 90% Bound for Linearised IBE Metric
$\hat{\nu}_{C.IBE}$: Upper 90% Bound for Linearised Constant-Scaled IBE Metric

Table 54: Hyslop et al (2000) Upper 90% Bound for Linearised Population and Individual Bioequivalence FDA Metric of AUC in Data Sets A through ZF

Data	$\hat{\sigma}_R$	$\hat{\nu}_{PBE}$	$\hat{\nu}_{C.PBE}$	$\hat{\sigma}_{WR}$	$\hat{\nu}_{IBE}$	$\hat{\nu}_{C.IBE}$
ZE3	0.339	-0.029	0.073	0.155	-0.010	-0.061
ZF	0.603	-0.460	-0.004	0.344	-0.195	-0.065

$\hat{\sigma}_R$: MoM Estimate for Total SD, Reference Formula
$\hat{\nu}_{PBE}$: Upper 90% Bound for Linearised PBE Metric
$\hat{\nu}_{C.PBE}$: Upper 90% Bound for Linearised Constant-Scaled PBE Metric
$\hat{\sigma}_{WR}$: MoM Estimate for Within-Subject SD, Reference Formula
$\hat{\nu}_{IBE}$: Upper 90% Bound for Linearised IBE Metric
$\hat{\nu}_{C.IBE}$: Upper 90% Bound for Linearised Constant-Scaled IBE Metric

Table 55: Hyslop et al (2000) Upper 90% Bound for Linearised Population and Individual Bioequivalence FDA Metric of Cmax in Data Sets A through ZF

Data	$\hat{\sigma}_R$	$\hat{\nu}_{PBE}$	$\hat{\nu}_{C.PBE}$	$\hat{\sigma}_{WR}$	$\hat{\nu}_{IBE}$	$\hat{\nu}_{C.IBE}$
A	0.497	-0.195	0.091	0.330	-0.091	0.016
B	1.003	-1.086	0.336	0.444	-0.361	-0.065
C1	0.273	-0.069	-0.028	0.239	-0.087	-0.083
C2	0.445	-0.266	-0.065	0.295	-0.078	-0.011
D	0.523	-0.235	0.080	0.273	0.039	0.094
E	0.226	0.018	0.028	0.209	0.012	-0.003
F	0.648	-0.428	0.028	0.380	0.107	0.291
G	0.573	0.100	0.494	0.328	0.500	0.633
H	0.343	-0.083	0.005	0.168	-0.006	-0.051
I1	0.461	-0.155	0.064	0.218	0.080	0.080
I2	0.714	-0.424	0.178	0.307	0.309	0.412
J	0.470	-0.150	0.089	0.281	-0.011	0.042
K1	0.733	-0.576	0.124	0.474	-0.388	-0.074
K2	0.256	-0.024	0.005	0.173	-0.021	-0.062
K3	0.581	-0.224	0.200	0.353	-0.117	0.025
L1	0.454	-0.218	0.010	0.300	-0.114	-0.043
L2	0.694	-0.478	0.144	0.421	-0.270	-0.042
M	0.481	-0.055	0.209	0.278	0.074	0.133
N1	0.204	0.056	0.051	0.123	0.103	0.038
N2	0.386	-0.144	-0.011	0.193	0.038	0.017
O1	0.336	-0.090	0.004	0.261	-0.031	0.001
O2	0.441	-0.084	0.136	0.318	-0.028	0.071
P	0.231	-0.014	-0.005	0.154	0.067	0.017
Q1	0.253	-0.064	-0.044	0.161	0.015	-0.032
Q2	0.303	-0.013	0.054	0.196	0.088	0.070
R	1.089	-1.158	0.524	0.390	-0.058	0.162
S	1.111	-1.554	0.218	0.516	-0.345	0.110
T	0.881	-0.729	0.381	0.436	-0.033	0.281
U	0.770	-0.408	0.410	0.508	-0.462	-0.082
V	0.677	-0.476	0.107	0.547	-0.440	-0.017
W1	0.355	-0.039	0.079	0.193	-0.015	-0.040
W2	0.354	-0.042	0.075	0.195	-0.021	-0.046
W3	0.424	-0.080	0.115	0.219	0.031	0.032
W4	0.462	-0.243	-0.017	0.237	-0.054	-0.044
W5	0.389	-0.167	-0.024	0.224	-0.064	-0.068
W6	0.475	-0.234	0.017	0.272	-0.108	-0.068
X	0.321	0.024	0.106	0.170	0.086	0.049
Y	0.390	-0.044	0.117	0.380	-0.149	0.008
ZA	0.584	-0.330	0.090	0.469	-0.351	-0.072
ZB	0.362	0.153	0.290	0.271	0.172	0.230
ZC1	0.460	-0.264	-0.029	0.337	-0.137	-0.015
ZC2	0.333	-0.151	-0.067	0.187	-0.046	-0.077
ZC3	0.395	-0.177	-0.028	0.198	-0.039	-0.062
ZD1	1.245	-1.593	0.481	0.470	-0.325	-0.012
ZD2	0.327	-0.100	-0.016	0.190	-0.071	-0.105
ZD3	0.164	-0.040	-0.072	0.115	-0.025	-0.100
ZD4	0.228	-0.048	-0.043	0.112	-0.007	-0.081
ZE1	0.773	-0.457	0.328	0.351	-0.007	0.146
ZE2	0.248	-0.016	0.005	0.121	0.013	-0.056

$\hat{\sigma}_R$: MoM Estimate for Total SD, Reference Formula
$\hat{\nu}_{PBE}$: Upper 90% Bound for Linearised PBE Metric
$\hat{\nu}_{C.PBE}$: Upper 90% Bound for Linearised Constant-Scaled PBE Metric
$\hat{\sigma}_{WR}$: MoM Estimate for Within-Subject SD, Reference Formula
$\hat{\nu}_{IBE}$: Upper 90% Bound for Linearised IBE Metric
$\hat{\nu}_{C.IBE}$: Upper 90% Bound for Linearised Constant-Scaled IBE Metric

Table 55: Hyslop et al (2000) Upper 90% Bound for Linearised Population and Individual Bioequivalence FDA Metric of Cmax in Data Sets A through ZF

Data	$\hat{\sigma}_R$	$\hat{\nu}_{PBE}$	$\hat{\nu}_{C.PBE}$	$\hat{\sigma}_{WR}$	$\hat{\nu}_{IBE}$	$\hat{\nu}_{C.IBE}$
ZE3	0.338	-0.040	0.059	0.144	-0.010	-0.069
ZF	0.714	-0.287	0.429	0.557	-0.223	0.306

$\hat{\sigma}_R$: MoM Estimate for Total SD, Reference Formula
$\hat{\nu}_{PBE}$: Upper 90% Bound for Linearised PBE Metric
$\hat{\nu}_{C.PBE}$: Upper 90% Bound for Linearised Constant-Scaled PBE Metric
$\hat{\sigma}_{WR}$: MoM Estimate for Within-Subject SD, Reference Formula
$\hat{\nu}_{IBE}$: Upper 90% Bound for Linearised IBE Metric
$\hat{\nu}_{C.IBE}$: Upper 90% Bound for Linearised Constant-Scaled IBE Metric

Table 56: Asymptotic Upper 90% Bound for Linearised Population Bioequivalence FDA Metric of AUC in Data Sets A through ZF using an Unconstrained (UN) and Constrained (CSH) REML Model

Data	$\hat{\sigma}_R$ UN	$\hat{\nu}_{PBE}$ UN	$\hat{\nu}_{C.PBE}$ UN	$\hat{\sigma}_R$ CSH	$\hat{\nu}_{PBE}$ CSH	$\hat{\nu}_{C.PBE}$ CSH
A	0.489	-0.265	-0.042	0.489	-0.260	-0.038
B	1.419	-2.577	0.100	1.419	-2.571	0.110
C1	0.257	-0.085	-0.071	0.257	-0.083	-0.069
C2	0.432	-0.237	-0.081	0.432	-0.237	-0.081
D	0.564	-0.340	-0.030	0.564	-0.340	-0.030
E	0.245	-0.040	-0.048	0.245	-0.040	-0.048
F	0.578	-0.266	-0.059	0.579	-0.240	-0.028
G	0.568	-0.262	-0.030	0.568	-0.262	-0.030
H	0.288	-0.066	-0.036	0.288	-0.066	-0.036
I1	0.408	-0.088	0.035	0.408	-0.088	0.035
I2	0.701	-0.362	0.111	0.701	-0.362	0.111
J	0.453	-0.181	-0.012	0.453	-0.181	-0.012
K1	0.517	-0.318	-0.054	0.517	-0.314	-0.050
K2	0.184	-0.022	-0.043	0.184	-0.022	-0.043
K3	0.286	-0.090	-0.053	0.287	-0.085	-0.048
L1	0.397	-0.136	0.006	0.397	-0.135	0.007
L2	0.646	-0.442	0.030	0.647	-0.427	0.044
M	0.524	-0.244	0.006	0.524	-0.244	0.006
N1	0.142	0.019	-0.022	0.142	0.020	-0.022
N2	0.398	-0.161	-0.041	0.398	-0.161	-0.041
O1	0.253	-0.037	-0.022	0.253	-0.037	-0.022
O2	0.416	-0.093	0.067	0.416	-0.093	0.067
P	0.266	-0.067	-0.056	0.266	-0.066	-0.055
Q1	0.179	0.009	-0.014	0.179	0.009	-0.014
Q2	0.290	-0.010	0.035	0.290	-0.010	0.035
R	1.450	-2.488	0.346	1.450	-2.488	0.346
S	1.398	-2.605	0.080	1.398	-2.605	0.082
T	1.195	-1.552	0.479	1.195	-1.552	0.479
U	0.639	-0.404	0.103	0.641	-0.388	0.111
V	0.584	-0.362	-0.012	0.584	-0.362	-0.012
W1	0.349	-0.113	-0.025	0.349	-0.111	-0.023
W2	0.360	-0.125	-0.027	0.360	-0.124	-0.025
W3	0.411	-0.140	0.017	0.411	-0.138	0.019
W4	0.476	-0.252	-0.042	0.476	-0.252	-0.042
W5	0.371	-0.150	-0.045	0.371	-0.150	-0.045
W6	0.627	-0.388	0.048	0.627	-0.388	0.048
X	0.332	-0.059	0.019	0.333	-0.051	0.025
Y	0.337	-0.072	0.005	0.337	-0.072	0.005
ZA	0.485	-0.274	-0.064	0.485	-0.274	-0.064
ZB	0.302	-0.068	-0.032	0.302	-0.068	-0.032
ZC1	0.465	-0.243	-0.032	0.465	-0.243	-0.032
ZC2	0.328	-0.144	-0.077	0.328	-0.144	-0.077
ZC3	0.399	-0.183	-0.054	0.399	-0.183	-0.054
ZD1	1.270	-1.894	-0.080	1.271	-1.850	-0.003
ZD2	0.429	-0.210	-0.068	0.429	-0.208	-0.063
ZD3	0.248	-0.070	-0.063	0.248	-0.069	-0.063
ZD4	0.279	-0.085	-0.058	0.279	-0.085	-0.058
ZE1	0.654	-0.368	0.121	0.654	-0.368	0.121
ZE2	0.250	-0.027	-0.008	0.250	-0.027	-0.008
ZE3	0.339	-0.076	0.019	0.339	-0.076	0.019

$\hat{\sigma}_R$: REML Estimate for Total SD, Reference Formula
$\hat{\nu}_{PBE}$: Upper 90% Bound for Linearised PBE Metric
$\hat{\nu}_{C.PBE}$: Upper 90% Bound for Linearised Constant-Scaled PBE Metric

Table 56: Asymptotic Upper 90% Bound for Linearised Population Bioequivalence FDA Metric of AUC in Data Sets A through ZF using an Unconstrained (UN) and Constrained (CSH) REML Model

Data	$\hat{\sigma}_R$ UN	$\hat{\nu}_{PBE}$ UN	$\hat{\nu}_{C.PBE}$ UN	$\hat{\sigma}_R$ CSH	$\hat{\nu}_{PBE}$ CSH	$\hat{\nu}_{C.PBE}$ CSH
ZF	0.606	-0.491	-0.070	0.606	-0.491	-0.069

$\hat{\sigma}_R$: REML Estimate for Total SD, Reference Formula
$\hat{\nu}_{PBE}$: Upper 90% Bound for Linearised PBE Metric
$\hat{\nu}_{C.PBE}$: Upper 90% Bound for Linearised Constant-Scaled PBE Metric

Table 57: Bootstrapped Upper Non-Parametric Percentile 90% Bound for Linearised Population and Individual Bioequivalence FDA Metric of AUC in Data Sets A through ZF using an Unconstrained (UN) REML Model

Data	$\hat{\sigma}_R$ UN	$\hat{\nu}_{PBE}$ UN	$\hat{\nu}_{C.PBE}$ UN	$\hat{\sigma}_{WR}$ UN	$\hat{\nu}_{IBE}$ UN	$\hat{\nu}_{C.IBE}$ UN
A	0.489	-0.216	-0.034	0.267	-0.038	-0.046
B	1.419	-2.457	0.079	0.420	-0.265	-0.046
C1	0.257	-0.072	-0.067	0.221	-0.044	-0.091
C2	0.432	-0.170	-0.085	0.278	-0.006	-0.054
D	0.564	-0.299	-0.033	0.170	-0.024	-0.082
E	0.245	-0.039	-0.043	0.081	0.008	-0.085
F	0.578	-0.182	-0.064	0.252	-0.043	-0.087
G	0.568	-0.099	-0.004	0.163	0.063	0.002
H	0.288	-0.046	-0.048	0.125	-0.001	-0.082
I1	0.408	-0.048	0.033	0.177	0.082	0.032
I2	0.701	-0.336	0.122	0.296	0.244	0.271
J	0.453	-0.161	0.000	0.227	-0.028	-0.020
K1	0.517	-0.235	-0.042	0.253	-0.104	-0.094
K2	0.184	-0.014	-0.041	0.131	-0.014	-0.086
K3	0.286	-0.079	-0.054	0.174	-0.041	-0.094
L1	0.397	-0.131	0.029	0.229	-0.024	-0.035
L2	0.646	-0.461	0.035	0.358	-0.159	-0.059
M	0.524	-0.164	0.036	0.221	-0.026	-0.073
N1	0.142	0.045	0.005	0.072	0.073	-0.016
N2	0.398	-0.151	-0.040	0.211	-0.019	-0.051
O1	0.253	-0.015	-0.009	0.165	0.008	-0.058
O2	0.416	-0.014	0.119	0.268	-0.017	0.010
P	0.266	-0.073	-0.058	0.115	-0.017	-0.099
Q1	0.179	0.003	-0.017	0.113	0.024	-0.059
Q2	0.290	-0.002	0.032	0.182	0.047	0.005
R	1.450	-2.484	0.294	0.345	0.062	0.190
S	1.398	-2.632	0.058	0.459	-0.114	0.196
T	1.195	-1.520	0.482	0.459	0.102	0.431
U	0.639	-0.369	0.117	0.455	-0.160	-0.034
V	0.584	-0.079	0.176	0.375	0.044	0.066
W1	0.349	-0.099	-0.018	0.136	-0.002	-0.070
W2	0.360	-0.108	-0.017	0.161	-0.007	-0.071
W3	0.411	-0.112	0.019	0.143	0.006	-0.057
W4	0.476	-0.214	-0.035	0.210	0.057	0.026
W5	0.371	-0.127	-0.042	0.195	0.013	-0.034
W6	0.627	-0.351	0.064	0.272	0.036	0.048
X	0.332	-0.037	0.010	0.187	-0.003	-0.061
Y	0.337	-0.043	-0.003	0.175	0.010	-0.048
ZA	0.485	-0.117	-0.017	0.352	-0.143	-0.095
ZB	0.302	-0.024	-0.019	0.202	0.027	-0.006
ZC1	0.465	-0.245	-0.029	0.253	-0.013	0.007
ZC2	0.328	-0.143	-0.080	0.182	-0.037	-0.074
ZC3	0.399	-0.174	-0.058	0.156	0.018	-0.037
ZD1	1.270	-1.599	-0.079	0.421	-0.150	0.033
ZD2	0.429	-0.205	-0.069	0.098	-0.002	-0.086
ZD3	0.248	-0.071	-0.064	0.116	-0.004	-0.081
ZD4	0.279	-0.089	-0.058	0.122	-0.006	-0.078

$\hat{\sigma}_R$: REML Estimate for Total SD, Reference Formula
$\hat{\nu}_{IBE}$: Upper 90% Bound for Linearised PBE Metric
$\hat{\nu}_{C.IBE}$: Upper 90% Bound for Linearised Constant-Scaled PBE Metric
$\hat{\sigma}_{WR}$: REML Estimate for Within SD, Reference Formula
$\hat{\nu}_{IBE}$: Upper 90% Bound for Linearised IBE Metric
$\hat{\nu}_{C.IBE}$: Upper 90% Bound for Linearised Constant-Scaled IBE Metric

Table 57: Bootstrapped Upper Non-Parametric Percentile 90% Bound for Linearised Population and Individual Bioequivalence FDA Metric of AUC in Data Sets A through ZF using an Unconstrained (UN) REML Model

Data	$\hat{\sigma}_R$ UN	$\hat{\nu}_{PBE}$ UN	$\hat{\nu}_{C.PBE}$ UN	$\hat{\sigma}_{WR}$ UN	$\hat{\nu}_{IBE}$ UN	$\hat{\nu}_{C.IBE}$ UN
ZE1	0.654	-0.383	0.097	0.258	0.035	0.041
ZE2	0.250	-0.032	-0.010	0.126	0.007	-0.060
ZE3	0.339	-0.059	0.013	0.155	-0.009	-0.074
ZF	0.606	-0.272	0.108	0.348	-0.009	0.020

$\hat{\sigma}_R$: REML Estimate for Total SD, Reference Formula
$\hat{\nu}_{IBE}$: Upper 90% Bound for Linearised PBE Metric
$\hat{\nu}_{C.IBE}$: Upper 90% Bound for Linearised Constant-Scaled PBE Metric
$\hat{\sigma}_{WR}$: REML Estimate for Within SD, Reference Formula
$\hat{\nu}_{IBE}$: Upper 90% Bound for Linearised IBE Metric
$\hat{\nu}_{C.IBE}$: Upper 90% Bound for Linearised Constant-Scaled IBE Metric

Table 58: Bootstrapped Upper Non-Parametric Percentile 90% Bound for Linearised Population and Individual Bioequivalence FDA Metric of Cmax in Data Sets A through ZF using an Unconstrained (UN) REML Model

Data	$\hat{\sigma}_R$ UN	$\hat{\nu}_{PBE}$ UN	$\hat{\nu}_{C.PBE}$ UN	$\hat{\sigma}_{WR}$ UN	$\hat{\nu}_{IBE}$ UN	$\hat{\nu}_{C.IBE}$ UN
A	0.491	-0.196	0.028	0.326	-0.066	-0.004
B	1.003	-1.258	0.076	0.444	-0.348	-0.066
C1	0.286	-0.081	-0.050	0.267	-0.117	-0.104
C2	0.445	-0.193	-0.087	0.295	-0.034	-0.038
D	0.525	-0.261	0.016	0.278	0.064	0.066
E	0.225	-0.009	-0.011	0.206	0.003	-0.034
F	0.645	-0.312	-0.050	0.370	0.028	0.115
G	0.576	0.151	0.363	0.328	0.440	0.521
H	0.338	-0.053	-0.054	0.166	-0.009	-0.076
I1	0.460	-0.116	0.032	0.215	0.083	0.052
I2	0.713	-0.330	0.148	0.305	0.290	0.335
J	0.476	-0.148	0.045	0.290	-0.047	0.010
K1	0.730	-0.539	0.006	0.467	-0.357	-0.098
K2	0.255	-0.039	-0.023	0.170	-0.025	-0.071
K3	0.579	-0.210	0.115	0.346	-0.075	0.022
L1	0.453	-0.234	-0.030	0.302	-0.124	-0.062
L2	0.689	-0.524	0.035	0.418	-0.212	-0.043
M	0.481	-0.066	0.097	0.276	0.078	0.079
N1	0.203	0.094	0.080	0.122	0.140	0.076
N2	0.383	-0.137	-0.039	0.194	0.025	-0.014
O1	0.336	-0.052	0.003	0.257	0.015	0.008
O2	0.441	-0.053	0.091	0.312	0.019	0.054
P	0.232	0.017	-0.005	0.150	0.057	0.000
Q1	0.247	-0.048	-0.044	0.159	0.022	-0.041
Q2	0.300	-0.003	0.048	0.199	0.090	0.059
R	1.085	-1.449	0.206	0.388	-0.035	0.151
S	1.110	-1.648	0.054	0.518	-0.252	0.145
T	0.882	-0.753	0.281	0.439	0.040	0.358
U	0.772	-0.522	0.215	0.513	-0.292	-0.046
V	0.669	-0.121	0.207	0.538	-0.031	0.209
W1	0.356	-0.067	0.022	0.195	-0.020	-0.045
W2	0.355	-0.075	0.018	0.196	-0.030	-0.056
W3	0.423	-0.121	0.047	0.217	0.041	0.027
W4	0.460	-0.219	-0.059	0.233	-0.044	-0.053
W5	0.388	-0.156	-0.058	0.222	-0.058	-0.075
W6	0.473	-0.240	-0.042	0.264	-0.096	-0.070
X	0.320	-0.003	0.043	0.168	0.084	0.013
Y	0.394	0.104	0.190	0.386	0.047	0.093
ZA	0.596	-0.214	0.054	0.477	-0.296	-0.087
ZB	0.381	0.096	0.212	0.315	0.171	0.199
ZC1	0.458	-0.263	-0.051	0.336	-0.119	-0.022
ZC2	0.331	-0.146	-0.085	0.190	-0.052	-0.083
ZC3	0.393	-0.177	-0.061	0.194	-0.044	-0.071
ZD1	1.253	-1.732	-0.070	0.511	-0.325	-0.006
ZD2	0.327	-0.120	-0.054	0.190	-0.056	-0.103
ZD3	0.169	-0.027	-0.070	0.129	-0.020	-0.102
ZD4	0.229	-0.050	-0.056	0.115	-0.003	-0.078

$\hat{\sigma}_R$: REML Estimate for Total SD, Reference Formula
$\hat{\nu}_{IBE}$: Upper 90% Bound for Linearised PBE Metric
$\hat{\nu}_{C.IBE}$: Upper 90% Bound for Linearised Constant-Scaled PBE Metric
$\hat{\sigma}_{WR}$: REML Estimate for Within SD, Reference Formula
$\hat{\nu}_{IBE}$: Upper 90% Bound for Linearised IBE Metric
$\hat{\nu}_{C.IBE}$: Upper 90% Bound for Linearised Constant-Scaled IBE Metric

Table 58: Bootstrapped Upper Non-Parametric Percentile 90% Bound for Linearised Population and Individual Bioequivalence FDA Metric of Cmax in Data Sets A through ZF using an Unconstrained (UN) REML Model

Data	$\hat{\sigma}_R$ UN	$\hat{\nu}_{PBE}$ UN	$\hat{\nu}_{C.PBE}$ UN	$\hat{\sigma}_{WR}$ UN	$\hat{\nu}_{IBE}$ UN	$\hat{\nu}_{C.IBE}$ UN
ZE1	0.773	-0.577	0.129	0.348	0.004	0.127
ZE2	0.248	-0.036	-0.021	0.120	0.010	-0.059
ZE3	0.339	-0.065	-0.002	0.145	-0.011	-0.075
ZF	0.711	-0.550	0.202	0.553	-0.065	0.345

$\hat{\sigma}_R$: REML Estimate for Total SD, Reference Formula
$\hat{\nu}_{IBE}$: Upper 90% Bound for Linearised PBE Metric
$\hat{\nu}_{C.IBE}$: Upper 90% Bound for Linearised Constant-Scaled PBE Metric
$\hat{\sigma}_{WR}$: REML Estimate for Within SD, Reference Formula
$\hat{\nu}_{IBE}$: Upper 90% Bound for Linearised IBE Metric
$\hat{\nu}_{C.IBE}$: Upper 90% Bound for Linearised Constant-Scaled IBE Metric

Table 59: Asymptotic Upper 90% Bound for Linearised Individual Bioequivalence FDA Metric of AUC in Data Sets A through ZF using an Unconstrained (UN) and Constrained (RIS) REML Model

Data	$\hat{\sigma}_{WR}$ UN	$\hat{\nu}_{IBE}$ UN	$\hat{\nu}_{C.IBE}$ UN	$\hat{\sigma}_{WR}$ RIS	$\hat{\nu}_{IBE}$ RIS	$\hat{\nu}_{C.IBE}$ RIS
A	0.267	-0.067	-0.059	0.265	-0.087	-0.069
B	0.420	-0.277	-0.048	0.408	-0.265	-0.041
C1	0.221	-0.068	-0.095	0.212	-0.076	-0.099
C2	0.278	-0.067	-0.047	0.286	-0.089	-0.060
D	0.170	-0.028	-0.082	0.170	-0.029	-0.083
E	0.081	0.006	-0.084	0.081	0.006	-0.084
F	0.252	-0.048	-0.088	0.225	-0.035	-0.074
G	0.163	0.061	0.002	0.166	0.057	0.000
H	0.125	-0.005	-0.083	0.124	-0.004	-0.082
I1	0.177	0.078	0.033	0.181	0.074	0.031
I2	0.296	0.221	0.275	0.312	0.190	0.258
J	0.227	0.011	-0.010	0.234	-0.001	-0.016
K1	0.253	-0.090	-0.088	0.245	-0.096	-0.092
K2	0.131	-0.015	-0.087	0.125	-0.011	-0.084
K3	0.174	-0.041	-0.093	0.159	-0.026	-0.082
L1	0.229	-0.035	-0.048	0.225	-0.041	-0.052
L2	0.358	-0.162	-0.057	0.333	-0.128	-0.032
M	0.221	-0.045	-0.082	0.218	-0.043	-0.080
N1	0.072	0.041	-0.048	0.072	0.040	-0.048
N2	0.211	-0.016	-0.047	0.218	-0.026	-0.052
O1	0.165	-0.007	-0.066	0.161	-0.004	-0.064
O2	0.268	0.006	0.019	0.260	0.020	0.030
P	0.115	-0.014	-0.096	0.108	-0.013	-0.096
Q1	0.113	0.018	-0.060	0.107	0.021	-0.058
Q2	0.182	0.050	0.009	0.183	0.048	0.008
R	0.345	0.049	0.182	0.346	0.048	0.182
S	0.459	-0.186	0.131	0.466	-0.238	0.100
T	0.459	-0.008	0.324	0.456	-0.001	0.327
U	0.455	-0.328	-0.096	0.411	-0.250	-0.052
V	0.375	-0.153	-0.056	0.368	-0.151	-0.054
W1	0.136	-0.006	-0.074	0.132	-0.002	-0.071
W2	0.161	-0.017	-0.074	0.157	-0.017	-0.073
W3	0.143	0.000	-0.065	0.138	0.004	-0.062
W4	0.210	0.043	0.025	0.214	0.038	0.023
W5	0.195	-0.007	-0.042	0.196	-0.011	-0.043
W6	0.272	0.004	0.035	0.272	0.004	0.035
X	0.187	-0.012	-0.064	0.172	0.001	-0.053
Y	0.175	-0.003	-0.060	0.171	0.000	-0.057
ZA	0.352	-0.145	-0.079	0.349	-0.162	-0.087
ZB	0.202	0.017	-0.024	0.205	0.011	-0.027
ZC1	0.253	-0.011	0.003	0.261	-0.027	-0.005
ZC2	0.182	-0.033	-0.076	0.194	-0.050	-0.085
ZC3	0.156	0.015	-0.040	0.163	0.007	-0.045
ZD1	0.421	-0.179	0.011	0.388	-0.081	0.093
ZD2	0.098	-0.003	-0.087	0.094	0.000	-0.084
ZD3	0.116	-0.005	-0.082	0.121	-0.009	-0.084
ZD4	0.122	-0.006	-0.081	0.124	-0.008	-0.082
ZE1	0.258	0.028	0.047	0.256	0.030	0.047
ZE2	0.126	0.011	-0.060	0.124	0.012	-0.060
ZE3	0.155	-0.007	-0.066	0.150	-0.003	-0.063
ZF	0.348	-0.179	-0.067	0.354	-0.232	-0.097

$\hat{\sigma}_{WR}$: REML Estimate for Within-Subject SD, Reference Formula
$\hat{\nu}_{IBE}$: Upper 90% Bound for Linearised IBE Metric
$\hat{\nu}_{C.IBE}$: Upper 90% Bound for Linearised Constant-Scaled IBE Metric

Table 60: Asymptotic Upper 90% Bound for Linearised Population Bioequivalence FDA Metric of Cmax in Data Sets A through ZF using an Unconstrained (UN) and Constrained (CSH) REML Model

Data	$\hat{\sigma}_R$ UN	$\hat{\nu}_{PBE}$ UN	$\hat{\nu}_{C.PBE}$ UN	$\hat{\sigma}_R$ CSH	$\hat{\nu}_{PBE}$ CSH	$\hat{\nu}_{C.PBE}$ CSH
A	0.491	-0.216	0.026	0.491	-0.215	0.027
B	1.003	-1.257	0.094	1.003	-1.251	0.099
C1	0.286	-0.089	-0.051	0.286	-0.088	-0.050
C2	0.445	-0.258	-0.089	0.445	-0.258	-0.089
D	0.525	-0.260	0.014	0.525	-0.260	0.014
E	0.225	-0.007	-0.014	0.225	-0.005	-0.012
F	0.645	-0.359	-0.072	0.645	-0.359	-0.072
G	0.576	0.023	0.340	0.576	0.023	0.340
H	0.338	-0.112	-0.054	0.338	-0.112	-0.054
I1	0.460	-0.162	0.007	0.460	-0.162	0.007
I2	0.713	-0.391	0.098	0.713	-0.391	0.098
J	0.476	-0.177	0.027	0.475	-0.169	0.034
K1	0.730	-0.635	-0.012	0.730	-0.625	-0.005
K2	0.255	-0.041	-0.019	0.255	-0.040	-0.018
K3	0.579	-0.286	0.095	0.579	-0.278	0.101
L1	0.453	-0.227	-0.029	0.453	-0.225	-0.028
L2	0.689	-0.515	0.030	0.690	-0.500	0.044
M	0.481	-0.132	0.077	0.481	-0.132	0.077
N1	0.203	0.041	0.030	0.203	0.041	0.030
N2	0.383	-0.145	-0.041	0.383	-0.145	-0.041
O1	0.336	-0.092	-0.021	0.336	-0.092	-0.021
O2	0.441	-0.118	0.066	0.441	-0.118	0.066
P	0.232	-0.012	-0.015	0.232	-0.012	-0.015
Q1	0.247	-0.051	-0.046	0.247	-0.051	-0.046
Q2	0.300	-0.005	0.046	0.300	-0.005	0.046
R	1.085	-1.340	0.244	1.085	-1.336	0.249
S	1.110	-1.646	0.049	1.110	-1.643	0.052
T	0.882	-0.803	0.276	0.882	-0.803	0.276
U	0.772	-0.582	0.198	0.772	-0.573	0.201
V	0.669	-0.484	0.005	0.670	-0.473	0.012
W1	0.356	-0.085	0.023	0.356	-0.082	0.025
W2	0.355	-0.089	0.017	0.355	-0.085	0.020
W3	0.423	-0.126	0.045	0.423	-0.126	0.045
W4	0.460	-0.263	-0.072	0.460	-0.263	-0.072
W5	0.388	-0.183	-0.063	0.388	-0.183	-0.063
W6	0.473	-0.265	-0.050	0.474	-0.264	-0.049
X	0.320	-0.023	0.035	0.320	-0.023	0.035
Y	0.394	-0.084	0.046	0.397	-0.085	0.046
ZA	0.596	-0.383	-0.016	0.596	-0.383	-0.016
ZB	0.381	0.039	0.169	0.381	0.039	0.169
ZC1	0.458	-0.259	-0.049	0.458	-0.258	-0.048
ZC2	0.331	-0.152	-0.082	0.331	-0.151	-0.081
ZC3	0.393	-0.188	-0.060	0.393	-0.187	-0.060
ZD1	1.253	-1.877	-0.064	1.254	-1.821	0.013
ZD2	0.327	-0.122	-0.052	0.327	-0.114	-0.046
ZD3	0.169	-0.043	-0.076	0.169	-0.042	-0.075
ZD4	0.229	-0.054	-0.056	0.229	-0.054	-0.056
ZE1	0.773	-0.551	0.146	0.773	-0.551	0.146
ZE2	0.248	-0.032	-0.017	0.248	-0.032	-0.017
ZE3	0.339	-0.086	0.005	0.339	-0.086	0.005

$\hat{\sigma}_R$: REML Estimate for Total SD, Reference Formula
$\hat{\nu}_{PBE}$: Upper 90% Bound for Linearised PBE Metric
$\hat{\nu}_{C.PBE}$: Upper 90% Bound for Linearised Constant-Scaled PBE Metric

Table 60: Asymptotic Upper 90% Bound for Linearised Population Bioequivalence FDA Metric of Cmax in Data Sets A through ZF using an Unconstrained (UN) and Constrained (CSH) REML Model

Data	$\hat{\sigma}_R$ UN	$\hat{\nu}_{PBE}$ UN	$\hat{\nu}_{C.PBE}$ UN	$\hat{\sigma}_R$ CSH	$\hat{\nu}_{PBE}$ CSH	$\hat{\nu}_{C.PBE}$ CSH
ZF	0.711	-0.303	0.365	0.711	-0.303	0.365

$\hat{\sigma}_R$: REML Estimate for Total SD, Reference Formula
$\hat{\nu}_{PBE}$: Upper 90% Bound for Linearised PBE Metric
$\hat{\nu}_{C.PBE}$: Upper 90% Bound for Linearised Constant-Scaled PBE Metric

Table 61: Asymptotic Upper 90% Bound for Linearised Individual Bioequivalence FDA Metric of Cmax in Data Sets A through ZF using an Unconstrained (UN) and Constrained (RIS) REML Model

Data	$\hat{\sigma}_{WR}$ UN	$\hat{\nu}_{IBE}$ UN	$\hat{\nu}_{C.IBE}$ UN	$\hat{\sigma}_{WR}$ RIS	$\hat{\nu}_{IBE}$ RIS	$\hat{\nu}_{C.IBE}$ RIS
A	0.326	-0.067	-0.002	0.324	-0.100	-0.020
B	0.444	-0.333	-0.064	0.432	-0.333	-0.067
C1	0.267	-0.105	-0.100	0.246	-0.086	-0.088
C2	0.295	-0.066	-0.031	0.307	-0.100	-0.050
D	0.278	0.030	0.063	0.281	0.022	0.059
E	0.206	0.009	-0.035	0.204	-0.009	-0.041
F	0.370	0.083	0.139	0.407	-0.007	0.080
G	0.328	0.408	0.492	0.337	0.395	0.490
H	0.166	-0.009	-0.071	0.169	-0.013	-0.074
I1	0.215	0.072	0.049	0.222	0.064	0.047
I2	0.305	0.283	0.351	0.321	0.253	0.337
J	0.290	0.006	0.034	0.293	-0.018	0.024
K1	0.467	-0.318	-0.073	0.450	-0.333	-0.081
K2	0.170	-0.016	-0.068	0.164	-0.017	-0.068
K3	0.346	-0.082	0.016	0.334	-0.072	0.026
L1	0.302	-0.095	-0.048	0.301	-0.125	-0.065
L2	0.418	-0.213	-0.032	0.395	-0.190	-0.012
M	0.276	0.053	0.069	0.274	0.053	0.069
N1	0.122	0.086	0.017	0.123	0.087	0.017
N2	0.194	0.025	-0.011	0.201	0.017	-0.015
O1	0.257	-0.012	-0.012	0.259	-0.021	-0.016
O2	0.312	0.001	0.052	0.303	0.016	0.064
P	0.150	0.064	0.003	0.154	0.059	0.001
Q1	0.159	0.016	-0.042	0.167	0.005	-0.048
Q2	0.199	0.090	0.063	0.206	0.080	0.058
R	0.388	-0.047	0.144	0.385	-0.050	0.143
S	0.518	-0.325	0.103	0.525	-0.392	0.065
T	0.439	-0.026	0.267	0.437	-0.022	0.269
U	0.513	-0.411	-0.085	0.486	-0.338	-0.045
V	0.538	-0.331	-0.028	0.514	-0.377	-0.051
W1	0.195	-0.014	-0.049	0.187	-0.008	-0.045
W2	0.196	-0.020	-0.056	0.186	-0.011	-0.049
W3	0.217	0.023	0.007	0.215	0.024	0.007
W4	0.233	-0.044	-0.054	0.243	-0.061	-0.062
W5	0.222	-0.053	-0.073	0.228	-0.065	-0.079
W6	0.264	-0.079	-0.066	0.267	-0.105	-0.082
X	0.168	0.060	0.008	0.168	0.060	0.008
Y	0.386	-0.112	-0.017	0.356	-0.115	-0.011
ZA	0.477	-0.246	-0.040	0.462	-0.241	-0.032
ZB	0.315	0.078	0.128	0.297	0.108	0.150
ZC1	0.336	-0.118	-0.026	0.348	-0.189	-0.065
ZC2	0.190	-0.044	-0.082	0.199	-0.067	-0.095
ZC3	0.194	-0.032	-0.067	0.201	-0.051	-0.078
ZD1	0.511	-0.347	-0.025	0.471	-0.221	0.073
ZD2	0.190	-0.061	-0.103	0.169	-0.038	-0.088
ZD3	0.129	-0.030	-0.103	0.125	-0.033	-0.105
ZD4	0.115	-0.006	-0.084	0.117	-0.008	-0.085
ZE1	0.348	0.006	0.117	0.349	0.003	0.116
ZE2	0.120	0.013	-0.061	0.119	0.014	-0.061
ZE3	0.145	-0.008	-0.072	0.140	-0.005	-0.070
ZF	0.553	-0.169	0.290	0.537	-0.125	0.316

$\hat{\sigma}_{WR}$: REML Estimate for Within-Subject SD, Reference Formula
$\hat{\nu}_{IBE}$: Upper 90% Bound for Linearised IBE Metric
$\hat{\nu}_{C.IBE}$: Upper 90% Bound for Linearised Constant-Scaled IBE Metric

Table 62: Simulation 1: Mean Bias (SE) in $\hat{\delta}$ for Sample Size 16 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Complete Data Set										
1	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
2	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
3	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
4	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
5	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
6	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
7	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
8	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
9	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
10	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003
11	-0.002	0.003	-0.002	0.003	-0.001	0.003	-0.002	0.003	-0.002	0.003
12	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004
13	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004
14	0.000	0.005	0.000	0.005	0.000	0.005	0.000	0.005	0.000	0.005
15	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004
16	-0.001	0.005	-0.001	0.005	-0.001	0.005	-0.001	0.005	-0.001	0.005
17	-0.003	0.005	-0.003	0.005	-0.003	0.005	-0.003	0.005	-0.003	0.005
18	-0.001	0.007	-0.001	0.007	-0.001	0.007	-0.001	0.007	-0.001	0.007
19	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
20	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
21	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
22	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
23	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
24	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
25	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
26	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
27	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
28	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003
29	-0.002	0.003	-0.002	0.003	-0.002	0.003	-0.002	0.003	-0.002	0.003
30	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004
31	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004
32	0.000	0.005	0.000	0.005	0.000	0.005	0.000	0.005	0.000	0.005
33	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004
34	-0.001	0.005	-0.001	0.005	-0.001	0.005	-0.001	0.005	-0.001	0.005
35	-0.003	0.005	-0.003	0.005	-0.003	0.005	-0.003	0.005	-0.003	0.005
36	-0.001	0.007	-0.001	0.007	-0.001	0.007	-0.001	0.007	-0.001	0.007
37	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
38	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
39	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
40	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
41	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
42	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
43	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
44	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
45	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
46	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003
47	-0.002	0.003	-0.002	0.003	-0.001	0.003	-0.002	0.003	-0.002	0.003
48	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004
49	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004
50	0.000	0.005	0.000	0.005	0.000	0.005	0.000	0.005	0.000	0.005
51	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004
52	-0.001	0.005	-0.001	0.005	-0.001	0.005	-0.001	0.005	-0.001	0.005
53	-0.003	0.005	-0.003	0.005	-0.003	0.005	-0.003	0.005	-0.003	0.005
54	-0.001	0.007	-0.001	0.007	-0.001	0.007	-0.001	0.007	-0.001	0.007

Table 62: Simulation 1: Mean Bias (SE) in $\hat{\delta}$ for Sample Size 16 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Substantial Missing Data										
1	0.062	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
2	0.062	0.001	-0.001	0.001	-0.002	0.001	-0.002	0.001	-0.001	0.001
3	0.062	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
4	0.061	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
5	0.062	0.002	-0.001	0.002	-0.002	0.002	-0.001	0.002	-0.001	0.002
6	0.060	0.002	-0.002	0.002	-0.003	0.002	-0.002	0.002	-0.002	0.002
7	0.062	0.003	-0.003	0.004	-0.004	0.004	-0.004	0.004	-0.003	0.003
8	0.060	0.004	-0.005	0.004	-0.005	0.004	-0.005	0.004	-0.005	0.004
9	0.061	0.004	-0.003	0.004	-0.004	0.004	-0.004	0.004	-0.004	0.004
10	0.059	0.005	-0.005	0.005	-0.006	0.005	-0.006	0.005	-0.006	0.005
11	0.061	0.005	-0.003	0.005	-0.005	0.005	-0.004	0.005	-0.005	0.005
12	0.057	0.006	-0.005	0.006	-0.007	0.006	-0.007	0.006	-0.006	0.006
13	0.061	0.006	-0.005	0.006	-0.007	0.006	-0.007	0.006	-0.006	0.006
14	0.058	0.007	-0.008	0.007	-0.010	0.007	-0.010	0.007	-0.009	0.007
15	0.060	0.007	-0.005	0.007	-0.007	0.007	-0.007	0.007	-0.007	0.007
16	0.056	0.008	-0.009	0.008	-0.010	0.008	-0.010	0.008	-0.010	0.008
17	0.059	0.008	-0.006	0.008	-0.007	0.008	-0.007	0.008	-0.007	0.008
18	0.053	0.011	-0.009	0.010	-0.012	0.010	-0.012	0.010	-0.011	0.010
19	0.062	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
20	0.062	0.001	-0.001	0.001	-0.002	0.001	-0.002	0.001	-0.001	0.001
21	0.062	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
22	0.061	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
23	0.062	0.002	-0.001	0.002	-0.002	0.002	-0.001	0.002	-0.001	0.002
24	0.060	0.002	-0.002	0.002	-0.003	0.002	-0.002	0.002	-0.002	0.002
25	0.062	0.003	-0.003	0.004	-0.004	0.004	-0.004	0.004	-0.003	0.003
26	0.060	0.004	-0.005	0.004	-0.005	0.004	-0.005	0.004	-0.005	0.004
27	0.061	0.004	-0.003	0.004	-0.004	0.004	-0.004	0.004	-0.004	0.004
28	0.059	0.005	-0.005	0.005	-0.006	0.005	-0.006	0.005	-0.006	0.005
29	0.061	0.005	-0.003	0.005	-0.005	0.005	-0.004	0.005	-0.005	0.005
30	0.057	0.006	-0.005	0.006	-0.007	0.006	-0.007	0.006	-0.006	0.006
31	0.061	0.006	-0.005	0.006	-0.007	0.006	-0.007	0.006	-0.006	0.006
32	0.058	0.007	-0.008	0.007	-0.010	0.007	-0.010	0.007	-0.009	0.007
33	0.060	0.007	-0.005	0.007	-0.007	0.007	-0.007	0.007	-0.007	0.007
34	0.056	0.008	-0.009	0.008	-0.010	0.008	-0.010	0.008	-0.010	0.008
35	0.059	0.008	-0.006	0.008	-0.007	0.008	-0.007	0.008	-0.007	0.008
36	0.053	0.011	-0.009	0.010	-0.013	0.010	-0.012	0.010	-0.011	0.010
37	0.062	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
38	0.062	0.001	-0.001	0.001	-0.002	0.001	-0.002	0.001	-0.001	0.001
39	0.062	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
40	0.061	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
41	0.062	0.002	-0.001	0.002	-0.002	0.002	-0.001	0.002	-0.001	0.002
42	0.060	0.002	-0.002	0.002	-0.003	0.002	-0.002	0.002	-0.002	0.002
43	0.062	0.003	-0.003	0.004	-0.004	0.004	-0.004	0.004	-0.003	0.003
44	0.060	0.004	-0.005	0.004	-0.006	0.004	-0.005	0.004	-0.005	0.004
45	0.061	0.004	-0.003	0.004	-0.005	0.004	-0.004	0.004	-0.004	0.004
46	0.059	0.005	-0.005	0.005	-0.006	0.005	-0.006	0.005	-0.006	0.005
47	0.061	0.005	-0.003	0.005	-0.004	0.005	-0.004	0.005	-0.005	0.005
48	0.057	0.006	-0.005	0.006	-0.007	0.006	-0.007	0.006	-0.006	0.006
49	0.061	0.006	-0.005	0.006	-0.007	0.006	-0.007	0.006	-0.006	0.006
50	0.058	0.007	-0.008	0.007	-0.010	0.007	-0.010	0.007	-0.009	0.007
51	0.060	0.007	-0.005	0.007	-0.007	0.007	-0.007	0.007	-0.007	0.007
52	0.056	0.008	-0.009	0.008	-0.010	0.008	-0.010	0.008	-0.010	0.008
53	0.059	0.008	-0.006	0.008	-0.008	0.008	-0.007	0.008	-0.007	0.008
54	0.053	0.011	-0.009	0.010	-0.013	0.010	-0.012	0.010	-0.011	0.010

Table 63: Simulation 1: Mean Bias (SE) in $\hat{\sigma}_D^2$ for Sample Size 16 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Complete Data Set										
1	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
2	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000
3	0.000	0.000	0.000	0.000	0.002	0.000	0.001	0.000	0.002	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.004	0.000	0.002	0.000	0.003	0.000
6	0.000	0.000	0.000	0.000	0.001	0.000	-0.004	0.000	0.000	0.000
7	-0.002	0.001	-0.002	0.001	0.013	0.001	0.009	0.000	0.011	0.001
8	-0.001	0.001	-0.001	0.001	0.005	0.001	-0.003	0.001	0.004	0.001
9	-0.002	0.001	-0.002	0.001	0.018	0.001	0.011	0.001	0.015	0.001
10	-0.001	0.002	-0.001	0.002	0.004	0.002	-0.012	0.002	0.002	0.002
11	-0.002	0.002	-0.003	0.002	0.026	0.001	0.015	0.001	0.021	0.001
12	0.000	0.003	0.000	0.003	0.005	0.003	-0.027	0.002	0.001	0.003
13	-0.006	0.003	-0.006	0.003	0.042	0.002	0.026	0.001	0.034	0.002
14	-0.003	0.004	-0.003	0.004	0.012	0.004	-0.017	0.003	0.008	0.004
15	-0.006	0.004	-0.007	0.004	0.051	0.002	0.031	0.002	0.041	0.002
16	-0.002	0.006	-0.002	0.006	0.011	0.005	-0.034	0.004	0.006	0.005
17	-0.006	0.005	-0.007	0.005	0.079	0.003	0.045	0.002	0.061	0.003
18	0.002	0.009	0.001	0.009	0.014	0.009	-0.087	0.007	0.003	0.009
19	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
20	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000
21	0.000	0.000	0.000	0.000	0.002	0.000	0.001	0.000	0.002	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	0.004	0.000	0.002	0.000	0.003	0.000
24	0.000	0.000	0.000	0.000	0.001	0.000	-0.004	0.000	0.000	0.000
25	-0.002	0.001	-0.002	0.001	0.013	0.001	0.009	0.000	0.011	0.001
26	-0.001	0.001	-0.001	0.001	0.005	0.001	-0.003	0.001	0.004	0.001
27	-0.002	0.001	-0.002	0.001	0.018	0.001	0.011	0.001	0.015	0.001
28	-0.001	0.002	-0.001	0.002	0.004	0.002	-0.012	0.002	0.002	0.002
29	-0.002	0.002	-0.003	0.002	0.026	0.001	0.015	0.001	0.021	0.001
30	0.000	0.003	0.000	0.003	0.005	0.003	-0.027	0.002	0.001	0.003
31	-0.006	0.003	-0.006	0.003	0.042	0.002	0.026	0.001	0.034	0.002
32	-0.003	0.004	-0.003	0.004	0.012	0.004	-0.017	0.003	0.008	0.004
33	-0.006	0.004	-0.007	0.004	0.051	0.002	0.031	0.002	0.041	0.002
34	-0.002	0.006	-0.002	0.006	0.011	0.005	-0.034	0.004	0.006	0.005
35	-0.006	0.005	-0.007	0.005	0.079	0.003	0.045	0.002	0.061	0.003
36	0.002	0.009	0.001	0.009	0.014	0.009	-0.087	0.007	0.003	0.009
37	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
38	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000
39	0.000	0.000	0.000	0.000	0.002	0.000	0.001	0.000	0.002	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.004	0.000	0.002	0.000	0.003	0.000
42	0.000	0.000	0.000	0.000	0.001	0.000	-0.004	0.000	0.000	0.000
43	-0.002	0.001	-0.002	0.001	0.013	0.001	0.009	0.000	0.011	0.001
44	-0.001	0.001	-0.001	0.001	0.005	0.001	-0.003	0.001	0.004	0.001
45	-0.002	0.001	-0.002	0.001	0.018	0.001	0.011	0.001	0.015	0.001
46	-0.001	0.002	-0.001	0.002	0.004	0.002	-0.012	0.002	0.002	0.002
47	-0.002	0.002	-0.003	0.002	0.026	0.001	0.015	0.001	0.021	0.001
48	0.000	0.003	0.000	0.003	0.005	0.003	-0.027	0.002	0.001	0.003
49	-0.006	0.003	-0.006	0.003	0.042	0.002	0.026	0.001	0.034	0.002
50	-0.003	0.004	-0.003	0.004	0.012	0.004	-0.017	0.003	0.008	0.004
51	-0.006	0.004	-0.007	0.004	0.051	0.002	0.031	0.002	0.041	0.002
52	-0.002	0.006	-0.002	0.006	0.011	0.005	-0.034	0.004	0.006	0.005
53	-0.006	0.005	-0.007	0.005	0.079	0.003	0.045	0.002	0.061	0.003
54	0.002	0.009	0.001	0.009	0.014	0.009	-0.087	0.007	0.003	0.009

Table 63: Simulation 1: Mean Bias (SE) in $\hat{\sigma}_D^2$ for Sample Size 16 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Substantial Missing Data										
1	0.363	0.001	0.000	0.000	0.003	0.000	0.002	0.000	0.002	0.000
2	0.364	0.002	0.000	0.000	0.003	0.000	0.001	0.000	0.002	0.000
3	0.364	0.002	0.000	0.000	0.005	0.000	0.003	0.000	0.003	0.000
4	0.365	0.002	0.000	0.000	0.003	0.000	0.000	0.000	0.002	0.000
5	0.366	0.002	0.001	0.000	0.008	0.000	0.004	0.000	0.005	0.000
6	0.367	0.003	0.001	0.001	0.004	0.001	-0.005	0.000	0.002	0.001
7	0.395	0.005	0.002	0.002	0.032	0.001	0.018	0.001	0.023	0.001
8	0.396	0.005	0.002	0.003	0.026	0.002	0.008	0.002	0.018	0.002
9	0.402	0.005	0.003	0.003	0.043	0.002	0.023	0.001	0.031	0.002
10	0.405	0.007	0.004	0.004	0.028	0.003	-0.004	0.003	0.018	0.003
11	0.413	0.006	0.006	0.004	0.057	0.002	0.029	0.002	0.041	0.002
12	0.416	0.009	0.005	0.006	0.033	0.005	-0.027	0.004	0.017	0.005
13	0.464	0.010	0.007	0.006	0.099	0.004	0.054	0.003	0.073	0.004
14	0.468	0.012	0.005	0.009	0.073	0.007	0.010	0.006	0.050	0.007
15	0.477	0.011	0.009	0.007	0.118	0.005	0.063	0.004	0.087	0.005
16	0.482	0.014	0.010	0.011	0.079	0.009	-0.011	0.008	0.050	0.009
17	0.515	0.014	0.017	0.011	0.169	0.007	0.083	0.005	0.123	0.007
18	0.521	0.020	0.015	0.017	0.092	0.015	-0.091	0.010	0.043	0.014
19	0.363	0.001	0.000	0.000	0.003	0.000	0.002	0.000	0.002	0.000
20	0.364	0.002	0.000	0.000	0.003	0.000	0.001	0.000	0.002	0.000
21	0.364	0.002	0.000	0.000	0.005	0.000	0.003	0.000	0.003	0.000
22	0.365	0.002	0.000	0.000	0.003	0.000	0.000	0.000	0.002	0.000
23	0.366	0.002	0.001	0.000	0.008	0.000	0.004	0.000	0.005	0.000
24	0.367	0.003	0.001	0.001	0.004	0.001	-0.005	0.000	0.002	0.001
25	0.395	0.005	0.002	0.002	0.032	0.001	0.018	0.001	0.023	0.001
26	0.396	0.005	0.002	0.003	0.026	0.002	0.008	0.002	0.018	0.002
27	0.402	0.005	0.003	0.003	0.042	0.002	0.023	0.001	0.031	0.002
28	0.405	0.007	0.004	0.004	0.028	0.003	-0.004	0.003	0.018	0.003
29	0.413	0.006	0.006	0.004	0.057	0.002	0.029	0.002	0.041	0.002
30	0.416	0.009	0.005	0.006	0.033	0.005	-0.027	0.004	0.017	0.005
31	0.464	0.010	0.007	0.006	0.099	0.004	0.054	0.003	0.073	0.004
32	0.468	0.012	0.005	0.009	0.072	0.007	0.010	0.006	0.050	0.007
33	0.477	0.011	0.009	0.007	0.118	0.005	0.063	0.004	0.087	0.005
34	0.482	0.014	0.010	0.011	0.079	0.009	-0.011	0.008	0.050	0.009
35	0.515	0.014	0.017	0.011	0.168	0.007	0.083	0.005	0.123	0.007
36	0.521	0.020	0.015	0.017	0.092	0.015	-0.091	0.010	0.043	0.014
37	0.363	0.001	0.000	0.000	0.003	0.000	0.002	0.000	0.002	0.000
38	0.364	0.002	0.000	0.000	0.003	0.000	0.001	0.000	0.002	0.000
39	0.364	0.002	0.000	0.000	0.005	0.000	0.003	0.000	0.003	0.000
40	0.365	0.002	0.000	0.000	0.003	0.000	0.000	0.000	0.002	0.000
41	0.366	0.002	0.001	0.000	0.008	0.000	0.004	0.000	0.005	0.000
42	0.367	0.003	0.001	0.001	0.004	0.001	-0.005	0.000	0.002	0.001
43	0.395	0.005	0.002	0.002	0.032	0.001	0.018	0.001	0.023	0.001
44	0.396	0.005	0.002	0.003	0.027	0.002	0.008	0.002	0.018	0.002
45	0.402	0.005	0.003	0.003	0.042	0.002	0.023	0.001	0.031	0.002
46	0.405	0.007	0.004	0.004	0.028	0.003	-0.004	0.003	0.018	0.003
47	0.413	0.006	0.006	0.004	0.057	0.002	0.029	0.002	0.041	0.002
48	0.416	0.009	0.005	0.006	0.033	0.005	-0.027	0.004	0.017	0.005
49	0.464	0.010	0.007	0.006	0.098	0.004	0.054	0.003	0.073	0.004
50	0.468	0.012	0.005	0.009	0.072	0.007	0.010	0.006	0.050	0.007
51	0.477	0.011	0.009	0.007	0.118	0.005	0.063	0.004	0.086	0.005
52	0.482	0.014	0.010	0.011	0.079	0.009	-0.011	0.008	0.050	0.009
53	0.515	0.014	0.017	0.011	0.169	0.007	0.083	0.005	0.123	0.007
54	0.521	0.020	0.015	0.017	0.092	0.015	-0.091	0.010	0.043	0.014

Table 64: Simulation 1: Mean Bias (SE) in $\hat{\sigma}^2_{WT}$ for Sample Size 16 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Complete Data Set										
1	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
2	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.001	0.001	0.001	0.001	-0.008	0.001	-0.008	0.001	-0.007	0.001
8	0.001	0.001	0.001	0.001	-0.003	0.001	-0.003	0.001	-0.002	0.001
9	0.001	0.001	0.001	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
10	0.001	0.001	0.001	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
11	0.001	0.001	0.001	0.001	-0.005	0.001	-0.005	0.001	-0.004	0.001
12	0.001	0.001	0.001	0.001	0.000	0.001	0.000	0.001	0.000	0.001
13	0.002	0.003	0.002	0.003	-0.021	0.002	-0.021	0.002	-0.016	0.002
14	0.002	0.003	0.002	0.003	-0.006	0.003	-0.006	0.003	-0.004	0.003
15	0.002	0.003	0.002	0.003	-0.018	0.002	-0.018	0.002	-0.014	0.003
16	0.002	0.003	0.002	0.003	-0.003	0.003	-0.003	0.003	-0.001	0.003
17	0.002	0.003	0.002	0.003	-0.012	0.003	-0.012	0.003	-0.010	0.003
18	0.002	0.003	0.002	0.003	0.000	0.003	0.000	0.003	0.001	0.003
19	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
20	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	0.001	0.001	0.001	0.001	-0.008	0.001	-0.008	0.001	-0.007	0.001
26	0.001	0.001	0.001	0.001	-0.003	0.001	-0.003	0.001	-0.002	0.001
27	0.001	0.001	0.001	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
28	0.001	0.001	0.001	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
29	0.001	0.001	0.001	0.001	-0.005	0.001	-0.005	0.001	-0.004	0.001
30	0.001	0.001	0.001	0.001	0.000	0.001	0.000	0.001	0.000	0.001
31	0.002	0.003	0.002	0.003	-0.021	0.002	-0.021	0.002	-0.016	0.002
32	0.002	0.003	0.002	0.003	-0.006	0.003	-0.006	0.003	-0.004	0.003
33	0.002	0.003	0.002	0.003	-0.018	0.002	-0.018	0.002	-0.014	0.003
34	0.002	0.003	0.002	0.003	-0.003	0.003	-0.003	0.003	-0.001	0.003
35	0.002	0.003	0.002	0.003	-0.012	0.003	-0.012	0.003	-0.010	0.003
36	0.002	0.003	0.002	0.003	0.000	0.003	0.000	0.003	0.001	0.003
37	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
38	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
42	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
43	0.001	0.001	0.001	0.001	-0.008	0.001	-0.008	0.001	-0.007	0.001
44	0.001	0.001	0.001	0.001	-0.003	0.001	-0.003	0.001	-0.002	0.001
45	0.001	0.001	0.001	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
46	0.001	0.001	0.001	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
47	0.001	0.001	0.001	0.001	-0.005	0.001	-0.005	0.001	-0.004	0.001
48	0.001	0.001	0.001	0.001	0.000	0.001	0.000	0.001	0.000	0.001
49	0.002	0.003	0.002	0.003	-0.021	0.002	-0.021	0.002	-0.016	0.002
50	0.002	0.003	0.002	0.003	-0.006	0.003	-0.006	0.003	-0.004	0.003
51	0.002	0.003	0.002	0.003	-0.018	0.002	-0.018	0.002	-0.014	0.003
52	0.002	0.003	0.002	0.003	-0.003	0.003	-0.003	0.003	-0.001	0.003
53	0.002	0.003	0.002	0.003	-0.012	0.003	-0.012	0.003	-0.010	0.003
54	0.002	0.003	0.002	0.003	0.000	0.003	0.000	0.003	0.001	0.003

Table 64: Simulation 1: Mean Bias (SE) in $\hat{\sigma}^2_{WT}$ for Sample Size 16 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Substantial Missing Data										
1	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.001	0.000
2	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
3	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
4	0.000	0.000	0.001	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
6	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.002	0.003	0.002	-0.013	0.001	-0.013	0.001	-0.010	0.001
8	0.000	0.002	0.005	0.002	-0.010	0.001	-0.010	0.001	-0.007	0.001
9	0.000	0.002	0.004	0.002	-0.011	0.001	-0.012	0.001	-0.008	0.001
10	0.000	0.002	0.005	0.002	-0.006	0.001	-0.006	0.001	-0.004	0.002
11	0.000	0.002	0.004	0.002	-0.009	0.001	-0.010	0.001	-0.007	0.001
12	0.000	0.002	0.005	0.002	-0.003	0.002	-0.004	0.002	-0.002	0.002
13	0.001	0.005	0.010	0.005	-0.035	0.003	-0.035	0.003	-0.026	0.004
14	0.001	0.005	0.014	0.005	-0.023	0.004	-0.023	0.004	-0.015	0.004
15	0.001	0.005	0.011	0.005	-0.031	0.004	-0.032	0.003	-0.024	0.004
16	0.001	0.005	0.013	0.005	-0.017	0.004	-0.018	0.004	-0.011	0.004
17	0.001	0.005	0.012	0.005	-0.024	0.004	-0.026	0.004	-0.020	0.004
18	0.001	0.005	0.015	0.006	-0.008	0.004	-0.009	0.004	-0.003	0.005
19	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.001	0.000
20	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
21	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
22	0.000	0.000	0.001	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
24	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.002	0.003	0.002	-0.013	0.001	-0.013	0.001	-0.010	0.001
26	0.000	0.002	0.005	0.002	-0.010	0.001	-0.010	0.001	-0.007	0.001
27	0.000	0.002	0.004	0.002	-0.011	0.001	-0.012	0.001	-0.008	0.001
28	0.000	0.002	0.005	0.002	-0.006	0.001	-0.006	0.001	-0.004	0.002
29	0.000	0.002	0.004	0.002	-0.009	0.001	-0.010	0.001	-0.007	0.001
30	0.000	0.002	0.005	0.002	-0.003	0.002	-0.004	0.002	-0.002	0.002
31	0.001	0.005	0.010	0.005	-0.035	0.003	-0.035	0.003	-0.026	0.004
32	0.001	0.005	0.014	0.005	-0.023	0.004	-0.023	0.004	-0.015	0.004
33	0.001	0.005	0.011	0.005	-0.031	0.004	-0.032	0.003	-0.024	0.004
34	0.001	0.005	0.013	0.005	-0.017	0.004	-0.018	0.004	-0.011	0.004
35	0.001	0.005	0.012	0.005	-0.025	0.004	-0.026	0.004	-0.020	0.004
36	0.001	0.005	0.015	0.006	-0.008	0.004	-0.009	0.004	-0.003	0.005
37	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.001	0.000
38	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
39	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
40	0.000	0.000	0.001	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
42	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
43	0.000	0.002	0.003	0.002	-0.013	0.001	-0.013	0.001	-0.010	0.001
44	0.000	0.002	0.005	0.002	-0.010	0.001	-0.010	0.001	-0.007	0.001
45	0.000	0.002	0.004	0.002	-0.011	0.001	-0.012	0.001	-0.008	0.001
46	0.000	0.002	0.005	0.002	-0.006	0.001	-0.006	0.001	-0.004	0.002
47	0.000	0.002	0.004	0.002	-0.009	0.001	-0.010	0.001	-0.007	0.001
48	0.000	0.002	0.005	0.002	-0.003	0.002	-0.004	0.002	-0.002	0.002
49	0.001	0.005	0.010	0.005	-0.035	0.003	-0.035	0.003	-0.026	0.004
50	0.001	0.005	0.014	0.005	-0.023	0.004	-0.023	0.004	-0.015	0.004
51	0.001	0.005	0.011	0.005	-0.031	0.004	-0.032	0.003	-0.023	0.004
52	0.001	0.005	0.013	0.005	-0.017	0.004	-0.018	0.004	-0.011	0.004
53	0.001	0.005	0.012	0.005	-0.025	0.004	-0.026	0.004	-0.020	0.004
54	0.001	0.005	0.015	0.006	-0.008	0.004	-0.009	0.004	-0.003	0.005

Table 65: Simulation 1: Mean Bias (SE) in $\hat{\sigma}^2_{WR}$ for Sample Size 16 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Complete Data Set										
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
7	-0.001	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
8	-0.001	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
9	-0.001	0.001	-0.001	0.001	-0.007	0.001	-0.007	0.001	-0.006	0.001
10	-0.001	0.001	-0.001	0.001	-0.002	0.001	-0.002	0.001	-0.001	0.001
11	-0.002	0.002	-0.001	0.002	-0.016	0.002	-0.016	0.002	-0.013	0.002
12	-0.002	0.002	-0.001	0.002	-0.004	0.002	-0.004	0.002	-0.001	0.002
13	-0.002	0.002	-0.001	0.002	-0.011	0.002	-0.011	0.002	-0.009	0.002
14	-0.002	0.002	-0.001	0.002	-0.004	0.002	-0.004	0.002	-0.003	0.002
15	-0.003	0.003	-0.002	0.003	-0.021	0.003	-0.021	0.003	-0.017	0.003
16	-0.003	0.003	-0.002	0.003	-0.006	0.003	-0.006	0.003	-0.004	0.003
17	-0.006	0.006	-0.004	0.006	-0.049	0.005	-0.050	0.005	-0.039	0.005
18	-0.006	0.006	-0.004	0.006	-0.011	0.006	-0.011	0.006	-0.003	0.006
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
25	-0.001	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
26	-0.001	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
27	-0.001	0.001	-0.001	0.001	-0.007	0.001	-0.007	0.001	-0.006	0.001
28	-0.001	0.001	-0.001	0.001	-0.002	0.001	-0.002	0.001	-0.001	0.001
29	-0.002	0.002	-0.001	0.002	-0.016	0.002	-0.016	0.002	-0.013	0.002
30	-0.002	0.002	-0.001	0.002	-0.004	0.002	-0.004	0.002	-0.001	0.002
31	-0.002	0.002	-0.001	0.002	-0.011	0.002	-0.011	0.002	-0.009	0.002
32	-0.002	0.002	-0.001	0.002	-0.004	0.002	-0.004	0.002	-0.003	0.002
33	-0.003	0.003	-0.002	0.003	-0.021	0.003	-0.021	0.003	-0.017	0.003
34	-0.003	0.003	-0.002	0.003	-0.006	0.003	-0.006	0.003	-0.004	0.003
35	-0.006	0.006	-0.004	0.006	-0.049	0.005	-0.050	0.005	-0.039	0.005
36	-0.006	0.006	-0.004	0.006	-0.011	0.006	-0.011	0.006	-0.003	0.006
37	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
42	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
43	-0.001	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
44	-0.001	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
45	-0.001	0.001	-0.001	0.001	-0.007	0.001	-0.007	0.001	-0.006	0.001
46	-0.001	0.001	-0.001	0.001	-0.002	0.001	-0.002	0.001	-0.001	0.001
47	-0.002	0.002	-0.001	0.002	-0.016	0.002	-0.016	0.002	-0.013	0.002
48	-0.002	0.002	-0.001	0.002	-0.004	0.002	-0.004	0.002	-0.001	0.002
49	-0.002	0.002	-0.001	0.002	-0.011	0.002	-0.011	0.002	-0.009	0.002
50	-0.002	0.002	-0.001	0.002	-0.004	0.002	-0.004	0.002	-0.003	0.002
51	-0.003	0.003	-0.002	0.003	-0.021	0.003	-0.021	0.003	-0.017	0.003
52	-0.003	0.003	-0.002	0.003	-0.006	0.003	-0.006	0.003	-0.004	0.003
53	-0.006	0.006	-0.004	0.006	-0.049	0.005	-0.050	0.005	-0.039	0.005
54	-0.006	0.006	-0.004	0.006	-0.011	0.006	-0.011	0.006	-0.003	0.006

Table 65: Simulation 1: Mean Bias (SE) in $\hat{\sigma}^2_{WR}$ for Sample Size 16 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Substantial Missing Data										
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
4	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	-0.003	0.000	-0.003	0.000	-0.002	0.000
6	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
7	-0.001	0.001	0.000	0.001	-0.003	0.001	-0.003	0.001	-0.002	0.001
8	-0.001	0.001	0.000	0.001	-0.002	0.001	-0.002	0.001	-0.002	0.001
9	-0.001	0.001	-0.001	0.001	-0.009	0.001	-0.009	0.001	-0.007	0.001
10	-0.001	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.003	0.001
11	-0.002	0.002	-0.002	0.002	-0.018	0.002	-0.018	0.002	-0.014	0.002
12	-0.002	0.002	-0.001	0.002	-0.009	0.002	-0.009	0.002	-0.004	0.002
13	-0.002	0.002	-0.001	0.003	-0.013	0.002	-0.013	0.002	-0.011	0.002
14	-0.002	0.002	0.000	0.003	-0.008	0.002	-0.008	0.002	-0.006	0.002
15	-0.003	0.004	-0.002	0.004	-0.024	0.003	-0.024	0.003	-0.019	0.003
16	-0.003	0.004	-0.001	0.004	-0.014	0.003	-0.014	0.003	-0.009	0.004
17	-0.007	0.008	-0.007	0.007	-0.055	0.006	-0.057	0.006	-0.042	0.006
18	-0.007	0.008	-0.005	0.007	-0.026	0.007	-0.027	0.007	-0.010	0.007
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
22	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	-0.003	0.000	-0.003	0.000	-0.002	0.000
24	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
25	-0.001	0.001	0.000	0.001	-0.003	0.001	-0.003	0.001	-0.002	0.001
26	-0.001	0.001	0.000	0.001	-0.002	0.001	-0.002	0.001	-0.002	0.001
27	-0.001	0.001	-0.001	0.001	-0.009	0.001	-0.009	0.001	-0.007	0.001
28	-0.001	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.003	0.001
29	-0.002	0.002	-0.002	0.002	-0.018	0.002	-0.018	0.002	-0.014	0.002
30	-0.002	0.002	-0.001	0.002	-0.009	0.002	-0.009	0.002	-0.004	0.002
31	-0.002	0.002	-0.001	0.003	-0.013	0.002	-0.013	0.002	-0.011	0.002
32	-0.002	0.002	0.000	0.003	-0.008	0.002	-0.008	0.002	-0.006	0.002
33	-0.003	0.004	-0.002	0.004	-0.024	0.003	-0.024	0.003	-0.019	0.003
34	-0.003	0.004	-0.001	0.004	-0.014	0.003	-0.014	0.003	-0.009	0.004
35	-0.007	0.008	-0.007	0.007	-0.055	0.006	-0.057	0.006	-0.042	0.006
36	-0.007	0.008	-0.005	0.007	-0.027	0.007	-0.027	0.007	-0.010	0.007
37	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
40	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	-0.003	0.000	-0.003	0.000	-0.002	0.000
42	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
43	-0.001	0.001	0.000	0.001	-0.003	0.001	-0.003	0.001	-0.002	0.001
44	-0.001	0.001	0.000	0.001	-0.002	0.001	-0.002	0.001	-0.002	0.001
45	-0.001	0.001	-0.001	0.001	-0.009	0.001	-0.009	0.001	-0.007	0.001
46	-0.001	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.003	0.001
47	-0.002	0.002	-0.002	0.002	-0.018	0.002	-0.018	0.002	-0.014	0.002
48	-0.002	0.002	-0.001	0.002	-0.009	0.002	-0.009	0.002	-0.004	0.002
49	-0.002	0.002	-0.001	0.003	-0.013	0.002	-0.013	0.002	-0.011	0.002
50	-0.002	0.002	0.000	0.003	-0.008	0.002	-0.008	0.002	-0.006	0.002
51	-0.003	0.004	-0.002	0.004	-0.024	0.003	-0.024	0.003	-0.019	0.003
52	-0.003	0.004	-0.001	0.004	-0.014	0.003	-0.014	0.003	-0.009	0.004
53	-0.007	0.008	-0.007	0.007	-0.055	0.006	-0.057	0.006	-0.042	0.006
54	-0.007	0.008	-0.005	0.007	-0.027	0.007	-0.027	0.007	-0.010	0.007

Table 66: Simulation 1: Percentage of ABE Failures for Sample Size 16 (1000 runs per simulation)

Sim	MoM	UN	CSH	FA0(2)	RIS
Complete Data Set					
1	0	0	0	0	0
2	0	1	1	1	0.3
3	0	1.3	1.3	1.3	0
4	0	17.1	17.1	17.1	1.1
5	0	4.6	4.8	4.5	0
6	0.1	30.1	30.2	30.1	10
7	9.6	27.1	28	27.9	13.9
8	25.8	58.3	58.8	58.7	38
9	23.7	37.7	42	42	32.9
10	59.7	79.1	79.8	79.8	70.7
11	47.4	56.2	60.5	60.2	55.6
12	85.8	93.3	93.5	93.5	91.3
13	77.3	79.2	85.5	85.5	88.4
14	94.3	95.3	97	97	97.1
15	87.1	87.2	93.2	93.2	95.1
16	98.2	98.3	99.5	99.5	99.5
17	96.9	96.9	99.4	99.4	99.9
18	99.9	99.9	100	100	100
19	95.1	95.7	97.4	97.4	97.2
20	95.6	96.8	97.3	97.3	96.8
21	94.2	94.4	96.4	96.4	96.7
22	95.2	97	97.2	97.2	96.4
23	94.3	94.4	96.4	96.4	96.7
24	94.9	97.1	97.4	97.3	96.4
25	94.3	95.1	96.9	96.9	96.8
26	95.9	97.3	97.4	97.4	96.7
27	94.2	94.4	96.4	96.4	96.8
28	95.7	97.2	97.5	97.5	97
29	94.4	94.6	96.4	96.4	96.9
30	97.3	98.6	98.8	98.8	98.5
31	95.3	95.8	98	98	97.9
32	98.9	99	99.5	99.5	99.4
33	96.2	96.3	98.8	98.8	99.2
34	99.6	99.7	99.9	99.9	99.9
35	98.3	98.3	99.9	99.9	100
36	100	100	100	100	100
37	100	100	100	100	100
38	100	100	100	100	100
39	100	100	100	100	100
40	100	100	100	100	100
41	100	100	100	100	100
42	100	100	100	100	100
43	100	100	100	100	100
44	100	100	100	100	100
45	100	100	100	100	100
46	100	100	100	100	100
47	100	100	100	100	100
48	100	100	100	100	100
49	100	100	100	100	100
50	100	100	100	100	100
51	100	100	100	100	100
52	100	100	100	100	100
53	100	100	100	100	100
54	100	100	100	100	100
Substantial Missing Data					

Table 66: Simulation 1: Percentage of ABE Failures for Sample Size 16 (1000 runs per simulation)

Sim	MoM	UN	CSH	FA0(2)	RIS
1	100	26.8	26.7	25.3	29.7
2	100	38.9	39	37.5	42.9
3	100	28.2	28.9	27.3	23.2
4	100	52.9	53.1	52.4	48.7
5	100	30.4	32.1	29.7	32.3
6	99.9	63	64.5	62.7	62.9
7	99.7	70.9	78.2	77.5	97.2
8	99.6	80	86.5	86.1	98.8
9	99.8	80.3	89.5	89.2	99.3
10	99.8	91.1	96	95.8	100
11	99.9	87.4	97.3	97.2	99.8
12	100	95.8	99.4	99.4	100
13	99.8	94.5	99.9	99.9	100
14	99.9	96.9	100	100	100
15	99.9	95.3	100	100	100
16	100	98.2	100	100	100
17	100	97	100	100	100
18	100	99.2	100	100	100
19	100	92.5	96.8	96.8	99.3
20	100	92.3	96.1	95.8	99.3
21	100	92.7	96.7	96.7	99.2
22	100	94.1	96.2	96.1	99
23	100	93.5	97.2	96.9	99.4
24	100	95.7	96.9	96.9	99
25	100	93.6	97.7	97.6	99.8
26	100	94.2	97.1	96.9	99.7
27	100	94.1	97.9	97.9	100
28	99.9	95.9	98.6	98.6	100
29	99.9	95.6	99.2	99.2	100
30	99.9	98	99.5	99.5	100
31	99.8	96.4	100	100	100
32	99.9	97.9	100	100	100
33	99.9	97.9	100	100	100
34	99.9	97	100	100	100
35	99.9	98.3	100	100	100
36	100	98.4	100	100	100
37	99.8	99.6	100	100	100
38	100	100	100	100	100
39	100	100	100	100	100
40	100	100	100	100	100
41	100	100	100	100	100
42	100	100	100	100	100
43	100	100	100	100	100
44	100	100	100	100	100
45	100	100	100	100	100
46	100	100	100	100	100
47	100	100	100	100	100
48	100	100	100	100	100
49	100	100	100	100	100
50	100	100	100	100	100
51	100	100	100	100	100
52	100	100	100	100	100
53	100	100	100	100	100
54	100	100	100	100	100

Table 67: Simulation 1: Mean Bias (SE) in $\hat{\delta}$ for Sample Size 24 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Complete Data Set										
1	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
2	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
3	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
4	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
5	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
6	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
7	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
8	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
9	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
10	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003
11	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
12	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003
13	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003
14	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004
15	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003
16	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004
17	-0.003	0.004	-0.003	0.004	-0.003	0.004	-0.003	0.004	-0.003	0.004
18	-0.001	0.006	-0.001	0.006	-0.001	0.006	-0.001	0.006	-0.001	0.006
19	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
20	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
21	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
22	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
23	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
24	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
25	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
26	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
27	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
28	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003
29	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
30	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003
31	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003
32	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004
33	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003
34	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004
35	-0.003	0.004	-0.003	0.004	-0.003	0.004	-0.003	0.004	-0.003	0.004
36	-0.001	0.006	-0.001	0.006	-0.001	0.006	-0.001	0.006	-0.001	0.006
37	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
38	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
39	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
40	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
41	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
42	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
43	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
44	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
45	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
46	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003
47	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
48	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003
49	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003
50	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004
51	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003
52	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.001	0.004
53	-0.003	0.004	-0.003	0.004	-0.003	0.004	-0.003	0.004	-0.003	0.004
54	-0.001	0.006	-0.001	0.006	-0.001	0.006	-0.001	0.006	-0.001	0.006

Table 67: Simulation 1: Mean Bias (SE) in $\hat{\delta}$ for Sample Size 24 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Substantial Missing Data										
1	0.041	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
2	0.041	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
3	0.041	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
4	0.041	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
5	0.040	0.001	-0.001	0.001	-0.001	0.001	-0.002	0.001	-0.002	0.001
6	0.040	0.001	-0.002	0.001	-0.002	0.001	-0.002	0.001	-0.002	0.001
7	0.039	0.002	-0.002	0.002	-0.003	0.002	-0.003	0.002	-0.003	0.002
8	0.039	0.003	-0.002	0.002	-0.003	0.002	-0.003	0.002	-0.003	0.002
9	0.038	0.003	-0.003	0.002	-0.003	0.002	-0.003	0.002	-0.004	0.002
10	0.039	0.003	-0.004	0.003	-0.004	0.003	-0.004	0.003	-0.004	0.003
11	0.038	0.003	-0.003	0.003	-0.004	0.003	-0.004	0.003	-0.005	0.003
12	0.038	0.004	-0.005	0.004	-0.005	0.004	-0.005	0.004	-0.005	0.004
13	0.037	0.004	-0.004	0.004	-0.005	0.004	-0.005	0.004	-0.006	0.004
14	0.037	0.005	-0.005	0.005	-0.005	0.005	-0.005	0.005	-0.006	0.005
15	0.036	0.004	-0.005	0.004	-0.006	0.004	-0.006	0.004	-0.007	0.004
16	0.036	0.005	-0.006	0.005	-0.006	0.005	-0.006	0.005	-0.007	0.005
17	0.035	0.005	-0.006	0.005	-0.007	0.005	-0.007	0.005	-0.008	0.005
18	0.035	0.007	-0.008	0.007	-0.009	0.007	-0.009	0.007	-0.009	0.007
19	0.041	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
20	0.041	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
21	0.041	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
22	0.041	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
23	0.040	0.001	-0.001	0.001	-0.002	0.001	-0.002	0.001	-0.002	0.001
24	0.040	0.001	-0.002	0.001	-0.002	0.001	-0.002	0.001	-0.002	0.001
25	0.039	0.002	-0.002	0.002	-0.003	0.002	-0.003	0.002	-0.003	0.002
26	0.039	0.003	-0.002	0.002	-0.003	0.002	-0.003	0.002	-0.003	0.002
27	0.038	0.003	-0.003	0.002	-0.003	0.002	-0.003	0.002	-0.004	0.002
28	0.039	0.003	-0.004	0.003	-0.004	0.003	-0.004	0.003	-0.004	0.003
29	0.038	0.003	-0.003	0.003	-0.004	0.003	-0.004	0.003	-0.005	0.003
30	0.038	0.004	-0.005	0.004	-0.005	0.004	-0.005	0.004	-0.005	0.004
31	0.037	0.004	-0.004	0.004	-0.005	0.004	-0.005	0.004	-0.006	0.004
32	0.037	0.005	-0.005	0.005	-0.005	0.005	-0.005	0.005	-0.006	0.005
33	0.036	0.004	-0.005	0.004	-0.006	0.004	-0.006	0.004	-0.007	0.004
34	0.036	0.005	-0.006	0.005	-0.006	0.005	-0.006	0.005	-0.007	0.005
35	0.035	0.005	-0.006	0.005	-0.007	0.005	-0.007	0.005	-0.008	0.005
36	0.035	0.007	-0.008	0.007	-0.009	0.007	-0.009	0.007	-0.009	0.007
37	0.041	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
38	0.041	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
39	0.041	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
40	0.041	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
41	0.040	0.001	-0.001	0.001	-0.001	0.001	-0.002	0.001	-0.002	0.001
42	0.040	0.001	-0.002	0.001	-0.002	0.001	-0.002	0.001	-0.002	0.001
43	0.039	0.002	-0.002	0.002	-0.003	0.002	-0.003	0.002	-0.003	0.002
44	0.039	0.003	-0.002	0.002	-0.003	0.002	-0.003	0.002	-0.003	0.002
45	0.038	0.003	-0.003	0.002	-0.003	0.002	-0.003	0.002	-0.004	0.002
46	0.039	0.003	-0.004	0.003	-0.004	0.003	-0.004	0.003	-0.004	0.003
47	0.038	0.003	-0.003	0.003	-0.004	0.003	-0.004	0.003	-0.005	0.003
48	0.038	0.004	-0.005	0.004	-0.005	0.004	-0.005	0.004	-0.005	0.004
49	0.037	0.004	-0.004	0.004	-0.005	0.004	-0.005	0.004	-0.006	0.004
50	0.037	0.005	-0.005	0.005	-0.005	0.005	-0.005	0.005	-0.006	0.005
51	0.036	0.004	-0.005	0.004	-0.006	0.004	-0.006	0.004	-0.007	0.004
52	0.036	0.005	-0.006	0.005	-0.006	0.005	-0.006	0.005	-0.007	0.005
53	0.035	0.005	-0.006	0.005	-0.007	0.005	-0.007	0.005	-0.008	0.005
54	0.035	0.007	-0.008	0.007	-0.009	0.007	-0.009	0.007	-0.009	0.007

Table 68: Simulation 1: Mean Bias (SE) in $\hat{\sigma}_D^2$ for Sample Size 24 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Complete Data Set										
1	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.002	0.000	0.001	0.000	0.001	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.003	0.000	0.002	0.000	0.002	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	-0.004	0.000	0.000	0.000
7	-0.001	0.001	-0.001	0.001	0.010	0.000	0.007	0.000	0.009	0.000
8	-0.001	0.001	-0.001	0.001	0.002	0.001	-0.003	0.001	0.002	0.001
9	-0.002	0.001	-0.002	0.001	0.014	0.001	0.009	0.001	0.012	0.001
10	-0.001	0.002	-0.001	0.002	0.001	0.002	-0.011	0.001	0.000	0.002
11	-0.002	0.001	-0.002	0.001	0.020	0.001	0.013	0.001	0.016	0.001
12	-0.001	0.002	-0.001	0.002	0.000	0.002	-0.024	0.002	-0.001	0.002
13	-0.004	0.002	-0.004	0.002	0.032	0.001	0.022	0.001	0.028	0.001
14	-0.002	0.003	-0.002	0.003	0.004	0.003	-0.016	0.003	0.002	0.003
15	-0.004	0.003	-0.005	0.003	0.039	0.002	0.026	0.001	0.033	0.002
16	-0.002	0.004	-0.002	0.004	0.002	0.004	-0.031	0.004	0.000	0.004
17	-0.006	0.004	-0.007	0.004	0.059	0.002	0.038	0.002	0.048	0.002
18	-0.002	0.007	-0.003	0.007	0.002	0.007	-0.077	0.006	-0.004	0.007
19	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.002	0.000	0.001	0.000	0.001	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	0.003	0.000	0.002	0.000	0.002	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	-0.004	0.000	0.000	0.000
25	-0.001	0.001	-0.001	0.001	0.010	0.000	0.007	0.000	0.009	0.000
26	-0.001	0.001	-0.001	0.001	0.002	0.001	-0.003	0.001	0.002	0.001
27	-0.002	0.001	-0.002	0.001	0.014	0.001	0.009	0.001	0.012	0.001
28	-0.001	0.002	-0.001	0.002	0.001	0.002	-0.011	0.001	0.000	0.002
29	-0.002	0.001	-0.002	0.001	0.020	0.001	0.013	0.001	0.016	0.001
30	-0.001	0.002	-0.001	0.002	0.000	0.002	-0.024	0.002	-0.001	0.002
31	-0.004	0.002	-0.004	0.002	0.032	0.001	0.022	0.001	0.028	0.001
32	-0.002	0.003	-0.002	0.003	0.004	0.003	-0.016	0.003	0.002	0.003
33	-0.004	0.003	-0.005	0.003	0.039	0.002	0.026	0.001	0.033	0.002
34	-0.002	0.004	-0.002	0.004	0.002	0.004	-0.031	0.004	0.000	0.004
35	-0.006	0.004	-0.007	0.004	0.059	0.002	0.038	0.002	0.048	0.002
36	-0.002	0.007	-0.003	0.007	0.002	0.007	-0.077	0.006	-0.004	0.007
37	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.002	0.000	0.001	0.000	0.001	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.003	0.000	0.002	0.000	0.002	0.000
42	0.000	0.000	0.000	0.000	0.000	0.000	-0.004	0.000	0.000	0.000
43	-0.001	0.001	-0.001	0.001	0.010	0.000	0.007	0.000	0.009	0.000
44	-0.001	0.001	-0.001	0.001	0.002	0.001	-0.003	0.001	0.002	0.001
45	-0.002	0.001	-0.002	0.001	0.014	0.001	0.009	0.001	0.012	0.001
46	-0.001	0.002	-0.001	0.002	0.001	0.002	-0.011	0.001	0.000	0.002
47	-0.002	0.001	-0.002	0.001	0.020	0.001	0.013	0.001	0.016	0.001
48	-0.001	0.002	-0.001	0.002	0.000	0.002	-0.024	0.002	-0.001	0.002
49	-0.004	0.002	-0.004	0.002	0.032	0.001	0.022	0.001	0.028	0.001
50	-0.002	0.003	-0.002	0.003	0.004	0.003	-0.016	0.003	0.002	0.003
51	-0.004	0.003	-0.005	0.003	0.039	0.002	0.026	0.001	0.033	0.002
52	-0.002	0.004	-0.002	0.004	0.002	0.004	-0.031	0.004	0.000	0.004
53	-0.006	0.004	-0.007	0.004	0.059	0.002	0.038	0.002	0.048	0.002
54	-0.002	0.007	-0.003	0.007	0.002	0.007	-0.077	0.006	-0.004	0.007

Table 68: Simulation 1: Mean Bias (SE) in $\hat{\sigma}_D^2$ for Sample Size 24 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Substantial Missing Data										
1	0.184	0.001	0.000	0.000	0.002	0.000	0.001	0.000	0.001	0.000
2	0.184	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000
3	0.184	0.001	0.000	0.000	0.002	0.000	0.002	0.000	0.002	0.000
4	0.185	0.001	0.000	0.000	0.001	0.000	-0.001	0.000	0.000	0.000
5	0.185	0.001	0.000	0.000	0.004	0.000	0.002	0.000	0.003	0.000
6	0.186	0.001	0.000	0.000	0.001	0.000	-0.004	0.000	0.000	0.000
7	0.199	0.003	-0.001	0.001	0.016	0.001	0.011	0.001	0.013	0.001
8	0.200	0.003	0.000	0.002	0.007	0.001	-0.001	0.001	0.005	0.001
9	0.203	0.003	-0.001	0.001	0.021	0.001	0.013	0.001	0.017	0.001
10	0.205	0.004	0.000	0.002	0.005	0.002	-0.010	0.002	0.003	0.002
11	0.209	0.004	-0.001	0.002	0.028	0.001	0.018	0.001	0.023	0.001
12	0.211	0.005	0.000	0.003	0.005	0.003	-0.026	0.003	0.002	0.003
13	0.235	0.005	-0.001	0.004	0.048	0.002	0.032	0.002	0.039	0.002
14	0.237	0.007	-0.001	0.005	0.017	0.004	-0.011	0.004	0.010	0.004
15	0.242	0.006	-0.001	0.004	0.059	0.003	0.038	0.002	0.047	0.003
16	0.245	0.008	-0.001	0.006	0.015	0.006	-0.028	0.005	0.008	0.006
17	0.261	0.008	-0.001	0.006	0.084	0.004	0.052	0.003	0.067	0.004
18	0.266	0.012	-0.001	0.010	0.014	0.009	-0.083	0.008	0.004	0.009
19	0.184	0.001	0.000	0.000	0.002	0.000	0.001	0.000	0.001	0.000
20	0.184	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000
21	0.184	0.001	0.000	0.000	0.002	0.000	0.001	0.000	0.002	0.000
22	0.185	0.001	0.000	0.000	0.001	0.000	-0.001	0.000	0.000	0.000
23	0.185	0.001	0.000	0.000	0.004	0.000	0.002	0.000	0.003	0.000
24	0.186	0.001	0.000	0.000	0.001	0.000	-0.004	0.000	0.000	0.000
25	0.199	0.003	-0.001	0.001	0.016	0.001	0.011	0.001	0.013	0.001
26	0.200	0.003	0.000	0.002	0.007	0.001	-0.001	0.001	0.005	0.001
27	0.203	0.003	-0.001	0.001	0.021	0.001	0.013	0.001	0.017	0.001
28	0.205	0.004	0.000	0.002	0.005	0.002	-0.010	0.002	0.003	0.002
29	0.209	0.004	-0.001	0.002	0.028	0.001	0.018	0.001	0.023	0.001
30	0.211	0.005	0.000	0.003	0.005	0.003	-0.026	0.003	0.002	0.003
31	0.235	0.005	-0.002	0.003	0.048	0.002	0.032	0.002	0.039	0.002
32	0.237	0.007	-0.001	0.005	0.017	0.004	-0.011	0.004	0.010	0.004
33	0.242	0.006	-0.002	0.004	0.059	0.003	0.037	0.002	0.047	0.003
34	0.245	0.008	-0.001	0.006	0.015	0.006	-0.028	0.005	0.008	0.006
35	0.261	0.008	-0.001	0.006	0.084	0.004	0.052	0.003	0.067	0.004
36	0.266	0.012	-0.001	0.010	0.014	0.009	-0.083	0.008	0.004	0.009
37	0.184	0.001	0.000	0.000	0.002	0.000	0.001	0.000	0.001	0.000
38	0.184	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000
39	0.184	0.001	0.000	0.000	0.002	0.000	0.001	0.000	0.002	0.000
40	0.185	0.001	0.000	0.000	0.001	0.000	-0.001	0.000	0.000	0.000
41	0.185	0.001	0.000	0.000	0.004	0.000	0.002	0.000	0.003	0.000
42	0.186	0.001	0.000	0.000	0.001	0.000	-0.004	0.000	0.000	0.000
43	0.199	0.003	-0.001	0.001	0.016	0.001	0.011	0.001	0.013	0.001
44	0.200	0.003	0.000	0.002	0.007	0.001	-0.001	0.001	0.005	0.001
45	0.203	0.003	-0.001	0.001	0.021	0.001	0.013	0.001	0.017	0.001
46	0.205	0.004	0.000	0.002	0.005	0.002	-0.010	0.002	0.003	0.002
47	0.209	0.004	-0.001	0.002	0.028	0.001	0.018	0.001	0.023	0.001
48	0.211	0.005	0.000	0.003	0.005	0.003	-0.026	0.003	0.002	0.003
49	0.235	0.005	-0.002	0.003	0.048	0.002	0.032	0.002	0.039	0.002
50	0.237	0.007	-0.001	0.005	0.017	0.004	-0.011	0.004	0.010	0.004
51	0.242	0.006	-0.001	0.004	0.059	0.003	0.038	0.002	0.047	0.003
52	0.245	0.008	-0.001	0.006	0.015	0.006	-0.028	0.005	0.008	0.006
53	0.261	0.008	-0.001	0.006	0.084	0.004	0.052	0.003	0.067	0.004
54	0.266	0.012	-0.001	0.010	0.014	0.009	-0.083	0.008	0.004	0.009

Table 69: Simulation 1: Mean Bias (SE) in $\hat{\sigma}^2_{WT}$ for Sample Size 24 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Complete Data Set										
1	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.001	0.001	0.001	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
8	0.001	0.001	0.001	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
9	0.001	0.001	0.001	0.001	-0.004	0.001	-0.004	0.001	-0.003	0.001
10	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
11	0.001	0.001	0.001	0.001	-0.003	0.001	-0.003	0.001	-0.003	0.001
12	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
13	0.003	0.002	0.003	0.002	-0.014	0.002	-0.014	0.002	-0.012	0.002
14	0.003	0.002	0.003	0.002	0.000	0.002	0.000	0.002	0.001	0.002
15	0.003	0.002	0.003	0.002	-0.012	0.002	-0.012	0.002	-0.010	0.002
16	0.003	0.002	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002
17	0.003	0.002	0.003	0.002	-0.007	0.002	-0.007	0.002	-0.007	0.002
18	0.003	0.002	0.003	0.002	0.003	0.002	0.002	0.002	0.003	0.002
19	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	0.001	0.001	0.001	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
26	0.001	0.001	0.001	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
27	0.001	0.001	0.001	0.001	-0.004	0.001	-0.004	0.001	-0.003	0.001
28	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
29	0.001	0.001	0.001	0.001	-0.003	0.001	-0.003	0.001	-0.003	0.001
30	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
31	0.003	0.002	0.003	0.002	-0.014	0.002	-0.014	0.002	-0.012	0.002
32	0.003	0.002	0.003	0.002	0.000	0.002	0.000	0.002	0.001	0.002
33	0.003	0.002	0.003	0.002	-0.012	0.002	-0.012	0.002	-0.010	0.002
34	0.003	0.002	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002
35	0.003	0.002	0.003	0.002	-0.007	0.002	-0.007	0.002	-0.007	0.002
36	0.003	0.002	0.003	0.002	0.003	0.002	0.002	0.002	0.003	0.002
37	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
42	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
43	0.001	0.001	0.001	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
44	0.001	0.001	0.001	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
45	0.001	0.001	0.001	0.001	-0.004	0.001	-0.004	0.001	-0.003	0.001
46	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
47	0.001	0.001	0.001	0.001	-0.003	0.001	-0.003	0.001	-0.003	0.001
48	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
49	0.003	0.002	0.003	0.002	-0.014	0.002	-0.014	0.002	-0.012	0.002
50	0.003	0.002	0.003	0.002	0.000	0.002	0.000	0.002	0.001	0.002
51	0.003	0.002	0.003	0.002	-0.012	0.002	-0.012	0.002	-0.010	0.002
52	0.003	0.002	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002
53	0.003	0.002	0.003	0.002	-0.007	0.002	-0.007	0.002	-0.007	0.002
54	0.003	0.002	0.003	0.002	0.003	0.002	0.002	0.002	0.003	0.002

Table 69: Simulation 1: Mean Bias (SE) in $\hat{\sigma}^2_{WT}$ for Sample Size 24 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Substantial Missing Data										
1	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
2	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.001	0.001	0.001	0.001	-0.008	0.001	-0.008	0.001	-0.006	0.001
8	0.001	0.001	0.001	0.001	-0.003	0.001	-0.003	0.001	-0.002	0.001
9	0.001	0.001	0.001	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
10	0.001	0.001	0.001	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
11	0.001	0.001	0.001	0.001	-0.005	0.001	-0.005	0.001	-0.004	0.001
12	0.001	0.001	0.001	0.001	0.000	0.001	0.000	0.001	0.000	0.001
13	0.002	0.003	0.003	0.003	-0.021	0.002	-0.021	0.002	-0.016	0.002
14	0.002	0.003	0.003	0.003	-0.006	0.003	-0.006	0.003	-0.003	0.003
15	0.002	0.003	0.003	0.003	-0.018	0.003	-0.018	0.003	-0.014	0.003
16	0.002	0.003	0.003	0.003	-0.003	0.003	-0.003	0.003	-0.001	0.003
17	0.002	0.003	0.003	0.003	-0.013	0.003	-0.013	0.003	-0.011	0.003
18	0.002	0.003	0.003	0.003	0.000	0.003	0.000	0.003	0.000	0.003
19	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
20	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	0.001	0.001	0.001	0.001	-0.008	0.001	-0.008	0.001	-0.006	0.001
26	0.001	0.001	0.001	0.001	-0.003	0.001	-0.003	0.001	-0.002	0.001
27	0.001	0.001	0.001	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
28	0.001	0.001	0.001	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
29	0.001	0.001	0.001	0.001	-0.005	0.001	-0.005	0.001	-0.004	0.001
30	0.001	0.001	0.001	0.001	0.000	0.001	0.000	0.001	0.000	0.001
31	0.002	0.003	0.004	0.003	-0.021	0.002	-0.021	0.002	-0.016	0.002
32	0.002	0.003	0.003	0.003	-0.006	0.003	-0.006	0.003	-0.003	0.003
33	0.002	0.003	0.003	0.003	-0.018	0.003	-0.018	0.003	-0.014	0.003
34	0.002	0.003	0.003	0.003	-0.003	0.003	-0.003	0.003	-0.001	0.003
35	0.002	0.003	0.003	0.003	-0.013	0.003	-0.013	0.003	-0.011	0.003
36	0.002	0.003	0.003	0.003	0.000	0.003	0.000	0.003	0.000	0.003
37	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
38	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
42	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
43	0.001	0.001	0.001	0.001	-0.008	0.001	-0.008	0.001	-0.006	0.001
44	0.001	0.001	0.001	0.001	-0.003	0.001	-0.003	0.001	-0.002	0.001
45	0.001	0.001	0.001	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
46	0.001	0.001	0.001	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
47	0.001	0.001	0.001	0.001	-0.005	0.001	-0.005	0.001	-0.004	0.001
48	0.001	0.001	0.001	0.001	0.000	0.001	0.000	0.001	0.000	0.001
49	0.002	0.003	0.004	0.003	-0.021	0.002	-0.021	0.002	-0.016	0.002
50	0.002	0.003	0.003	0.003	-0.006	0.003	-0.006	0.003	-0.003	0.003
51	0.002	0.003	0.003	0.003	-0.018	0.003	-0.018	0.003	-0.014	0.003
52	0.002	0.003	0.003	0.003	-0.003	0.003	-0.003	0.003	-0.001	0.003
53	0.002	0.003	0.003	0.003	-0.013	0.003	-0.013	0.003	-0.011	0.003
54	0.002	0.003	0.003	0.003	0.000	0.003	0.000	0.003	0.000	0.003

Table 70: Simulation 1: Mean Bias (SE) in $\hat{\sigma}^2_{WR}$ for Sample Size 24 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Complete Data Set										
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.001	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.004	0.001
10	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
11	0.000	0.001	0.000	0.001	-0.011	0.001	-0.011	0.001	-0.009	0.001
12	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.001	0.002
13	0.000	0.001	0.000	0.002	-0.007	0.001	-0.007	0.001	-0.006	0.001
14	0.000	0.001	0.000	0.002	-0.001	0.001	-0.001	0.001	0.000	0.001
15	0.000	0.002	0.001	0.002	-0.014	0.002	-0.014	0.002	-0.012	0.002
16	0.000	0.002	0.001	0.002	-0.001	0.002	-0.001	0.002	0.000	0.002
17	0.000	0.005	0.001	0.005	-0.034	0.004	-0.034	0.004	-0.027	0.004
18	0.000	0.005	0.001	0.005	-0.001	0.004	-0.001	0.004	0.003	0.005
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.001	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
26	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
27	0.000	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.004	0.001
28	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
29	0.000	0.001	0.000	0.001	-0.011	0.001	-0.011	0.001	-0.009	0.001
30	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.001	0.002
31	0.000	0.001	0.000	0.002	-0.007	0.001	-0.007	0.001	-0.006	0.001
32	0.000	0.001	0.000	0.002	-0.001	0.001	-0.001	0.001	0.000	0.001
33	0.000	0.002	0.001	0.002	-0.014	0.002	-0.014	0.002	-0.012	0.002
34	0.000	0.002	0.001	0.002	-0.001	0.002	-0.001	0.002	0.000	0.002
35	0.000	0.005	0.001	0.005	-0.034	0.004	-0.034	0.004	-0.027	0.004
36	0.000	0.005	0.001	0.005	-0.001	0.004	-0.001	0.004	0.003	0.005
37	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.001	0.000
42	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
43	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
44	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
45	0.000	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.004	0.001
46	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
47	0.000	0.001	0.000	0.001	-0.011	0.001	-0.011	0.001	-0.009	0.001
48	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.001	0.002
49	0.000	0.001	0.000	0.002	-0.007	0.001	-0.007	0.001	-0.006	0.001
50	0.000	0.001	0.000	0.002	-0.001	0.001	-0.001	0.001	0.000	0.001
51	0.000	0.002	0.001	0.002	-0.014	0.002	-0.014	0.002	-0.012	0.002
52	0.000	0.002	0.001	0.002	-0.001	0.002	-0.001	0.002	0.000	0.002
53	0.000	0.005	0.001	0.005	-0.034	0.004	-0.034	0.004	-0.027	0.004
54	0.000	0.005	0.001	0.005	-0.001	0.004	-0.001	0.004	0.003	0.005

Table 70: Simulation 1: Mean Bias (SE) in $\hat{\sigma}^2_{WR}$ for Sample Size 24 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Substantial Missing Data										
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.001	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.001	0.000
8	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
9	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
10	0.000	0.001	0.000	0.001	-0.002	0.001	-0.002	0.001	-0.001	0.001
11	0.000	0.002	-0.001	0.002	-0.013	0.001	-0.012	0.001	-0.010	0.001
12	0.000	0.002	0.000	0.002	-0.002	0.002	-0.002	0.002	-0.001	0.002
13	0.000	0.002	-0.001	0.002	-0.008	0.002	-0.008	0.002	-0.007	0.002
14	0.000	0.002	0.000	0.002	-0.003	0.002	-0.003	0.002	-0.002	0.002
15	-0.001	0.003	-0.001	0.003	-0.016	0.002	-0.016	0.002	-0.014	0.002
16	-0.001	0.003	0.000	0.003	-0.004	0.003	-0.004	0.003	-0.003	0.003
17	-0.001	0.005	-0.002	0.005	-0.039	0.004	-0.039	0.004	-0.031	0.004
18	-0.001	0.005	-0.001	0.005	-0.007	0.005	-0.007	0.005	-0.002	0.005
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.001	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.001	0.000
26	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
27	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
28	0.000	0.001	0.000	0.001	-0.002	0.001	-0.002	0.001	-0.001	0.001
29	0.000	0.002	-0.001	0.002	-0.013	0.001	-0.012	0.001	-0.010	0.001
30	0.000	0.002	0.000	0.002	-0.002	0.002	-0.002	0.002	-0.001	0.002
31	0.000	0.002	-0.001	0.002	-0.008	0.002	-0.008	0.002	-0.007	0.002
32	0.000	0.002	0.000	0.002	-0.003	0.002	-0.003	0.002	-0.002	0.002
33	-0.001	0.003	-0.001	0.003	-0.016	0.002	-0.016	0.002	-0.014	0.002
34	-0.001	0.003	0.000	0.003	-0.004	0.003	-0.004	0.003	-0.003	0.003
35	-0.001	0.005	-0.002	0.005	-0.039	0.004	-0.039	0.004	-0.031	0.004
36	-0.001	0.005	-0.001	0.005	-0.007	0.005	-0.007	0.005	-0.002	0.005
37	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.001	0.000
42	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
43	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.001	0.000
44	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
45	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
46	0.000	0.001	0.000	0.001	-0.002	0.001	-0.002	0.001	-0.001	0.001
47	0.000	0.002	-0.001	0.002	-0.013	0.001	-0.012	0.001	-0.010	0.001
48	0.000	0.002	0.000	0.002	-0.002	0.002	-0.002	0.002	-0.001	0.002
49	0.000	0.002	-0.001	0.002	-0.008	0.002	-0.008	0.002	-0.007	0.002
50	0.000	0.002	0.000	0.002	-0.003	0.002	-0.003	0.002	-0.002	0.002
51	-0.001	0.003	-0.001	0.003	-0.016	0.002	-0.016	0.002	-0.014	0.002
52	-0.001	0.003	0.000	0.003	-0.004	0.003	-0.004	0.003	-0.003	0.003
53	-0.001	0.005	-0.002	0.005	-0.039	0.004	-0.039	0.004	-0.031	0.004
54	-0.001	0.005	-0.001	0.005	-0.007	0.005	-0.007	0.005	-0.002	0.005

Table 71: Simulation 1: Percentage of ABE Failures for Sample Size 24 (1000 runs per simulation)

Sim	MoM	UN	CSH	FA0(2)	RIS
Complete Data Set					
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	2.2	2.2	2.2	0
5	0	0.1	0.1	0.1	0
6	0	8.7	8.9	8.7	0.5
7	1.2	8.8	8.8	8.7	1.6
8	7.1	30.3	30.4	30.4	11.1
9	5.2	12.9	14	14.1	7.6
10	27.7	53.5	53.7	53.7	37.6
11	16.4	23.7	25.3	25.2	20.7
12	61.4	80.3	80.3	80.3	70.6
13	48.5	57.2	61.4	61.4	56.4
14	80.6	88.7	89.3	89.3	85.9
15	63.5	69.3	74	74	73.2
16	94	96.5	97.1	97.1	96.2
17	87.2	88	93.9	93.9	94.8
18	99.7	99.7	99.8	99.8	99.8
19	95.1	95.7	96.8	96.8	96.5
20	95.2	96	96.5	96.5	96.3
21	94.4	95.2	96.6	96.6	96.2
22	95.1	96.6	96.9	96.9	96.3
23	95	95.4	96.4	96.4	96.4
24	94.8	97.1	97.2	97.2	96.3
25	94.6	95.1	96.3	96.3	96.2
26	95.1	96.9	97.1	97.1	96.3
27	94.4	95.2	96.6	96.6	96.2
28	95.2	96.7	97	97	96.4
29	94.4	95	96.5	96.5	96.2
30	95.7	97.8	98	98	96.8
31	94.7	95.5	96.3	96.3	96.1
32	97.2	98.3	98.4	98.4	97.9
33	95	95.8	97.3	97.3	97
34	98.3	98.9	99.2	99.2	98.9
35	96.9	97.2	98.9	98.9	99
36	99.8	99.8	100	100	99.9
37	100	100	100	100	100
38	100	100	100	100	100
39	100	100	100	100	100
40	100	100	100	100	100
41	100	100	100	100	100
42	100	100	100	100	100
43	100	100	100	100	100
44	100	100	100	100	100
45	100	100	100	100	100
46	100	100	100	100	100
47	100	100	100	100	100
48	100	100	100	100	100
49	100	100	100	100	100
50	100	100	100	100	100
51	100	100	100	100	100
52	100	100	100	100	100
53	100	100	100	100	100
54	100	100	100	100	100
Substantial Missing Data					

Table 71: Simulation 1: Percentage of ABE Failures for Sample Size 24 (1000 runs per simulation)

Sim	MoM	UN	CSH	FA0(2)	RIS
1	12.6	0.3	0.4	0.3	0.3
2	16	1.5	1.5	1.5	0.9
3	18.3	1.8	1.8	1.8	0.3
4	27.4	13.2	13.2	13.3	3.7
5	25.6	3.5	3.6	3.5	0.5
6	37.5	25.9	26.2	25.9	12.4
7	59.3	28	29.7	28.8	21.8
8	71.5	54.6	55.9	54.9	45.4
9	69.4	39.3	42.6	42.3	36.7
10	84.2	78.5	78.7	78.2	72.7
11	81	55.9	60.4	59.4	59.3
12	94.2	92.6	93	92.9	92.3
13	92.1	80.9	87.8	87.3	92
14	97.4	95.9	97.5	97	99.1
15	94.5	88.7	95.7	95.5	98.1
16	98.7	98.7	99.8	99.7	100
17	97.9	96.9	100	100	100
18	99.9	99.8	100	100	100
19	100	94.9	96.3	96.1	96.9
20	100	95.8	96.8	96.4	96.5
21	100	95.9	97.3	97.2	96.9
22	100	96.7	96.9	96.9	95.8
23	100	96.5	97.1	97	96.8
24	100	97.1	97.2	97.1	96.3
25	99.9	95.2	96.6	96.4	96.9
26	99.5	96.2	96.4	96.3	96.1
27	99.8	95.9	97.3	97.2	96.9
28	99.3	96.8	97	97	96.3
29	99.5	96.2	97.1	96.9	96.9
30	98.7	98.2	98.4	98.4	97.9
31	99.2	96.7	98.4	98.2	98.8
32	99.3	98.9	99.1	99	99
33	99.5	97.4	99.3	99.3	99.6
34	99.9	99.3	99.7	99.7	99.8
35	99.7	99.1	99.9	99.9	100
36	100	100	100	100	100
37	100	100	100	100	100
38	100	100	100	100	100
39	100	100	100	100	100
40	100	100	100	100	100
41	100	100	100	100	100
42	100	100	100	100	100
43	100	100	100	100	100
44	100	100	100	100	100
45	100	100	100	100	100
46	100	100	100	100	100
47	100	100	100	100	100
48	100	100	100	100	100
49	100	100	100	100	100
50	100	100	100	100	100
51	100	100	100	100	100
52	100	100	100	100	100
53	100	100	100	100	100
54	100	100	100	100	100

Table 72: Simulation 1: Mean Bias (SE) in $\hat{\delta}$ for Sample Size 34 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Complete Data Set										
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
4	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
5	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
6	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
7	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
8	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
9	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
10	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
11	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
12	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003
13	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
14	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003
15	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003
16	0.000	0.004	0.000	0.004	0.000	0.004	0.000	0.004	0.000	0.004
17	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003
18	0.000	0.005	0.000	0.005	0.000	0.005	0.000	0.005	0.000	0.005
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
22	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
23	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
24	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
25	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
26	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
27	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
28	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
29	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
30	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003
31	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
32	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003
33	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003
34	0.000	0.004	0.000	0.004	0.000	0.004	0.000	0.004	0.000	0.004
35	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003
36	0.000	0.005	0.000	0.005	0.000	0.005	0.000	0.005	0.000	0.005
37	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
40	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
41	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
42	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
43	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
44	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
45	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
46	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
47	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
48	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003
49	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
50	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003
51	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003
52	0.000	0.004	0.000	0.004	0.000	0.004	0.000	0.004	0.000	0.004
53	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003
54	0.000	0.005	0.000	0.005	0.000	0.005	0.000	0.005	0.000	0.005

Table 72: Simulation 1: Mean Bias (SE) in $\hat{\delta}$ for Sample Size 34 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Substantial Missing Data										
1	0.029	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
2	0.029	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
3	0.029	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
4	0.029	0.001	-0.001	0.001	0.000	0.001	0.000	0.001	0.000	0.001
5	0.029	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
6	0.029	0.001	0.000	0.001	0.000	0.001	0.000	0.001	-0.001	0.001
7	0.027	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
8	0.028	0.002	-0.002	0.002	-0.001	0.002	-0.001	0.002	-0.002	0.002
9	0.027	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
10	0.028	0.002	-0.002	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
11	0.027	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
12	0.028	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.002	0.003
13	0.026	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003
14	0.027	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004
15	0.026	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003
16	0.027	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004
17	0.026	0.004	-0.003	0.004	-0.003	0.004	-0.003	0.004	-0.003	0.004
18	0.028	0.005	-0.002	0.005	-0.002	0.005	-0.002	0.005	-0.003	0.005
19	0.029	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
20	0.029	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
21	0.029	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
22	0.029	0.001	-0.001	0.001	0.000	0.001	0.000	0.001	0.000	0.001
23	0.029	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
24	0.029	0.001	0.000	0.001	0.000	0.001	0.000	0.001	-0.001	0.001
25	0.027	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
26	0.028	0.002	-0.002	0.002	-0.001	0.002	-0.001	0.002	-0.002	0.002
27	0.027	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
28	0.028	0.002	-0.002	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
29	0.027	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
30	0.028	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.002	0.003
31	0.026	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003
32	0.027	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004
33	0.026	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003
34	0.027	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004
35	0.026	0.004	-0.003	0.004	-0.003	0.004	-0.003	0.004	-0.003	0.004
36	0.028	0.005	-0.002	0.005	-0.002	0.005	-0.002	0.005	-0.003	0.005
37	0.029	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
38	0.029	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
39	0.029	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
40	0.029	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
41	0.029	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
42	0.029	0.001	0.000	0.001	0.000	0.001	0.000	0.001	-0.001	0.001
43	0.027	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
44	0.028	0.002	-0.002	0.002	-0.001	0.002	-0.001	0.002	-0.002	0.002
45	0.027	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
46	0.028	0.002	-0.002	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
47	0.027	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002	-0.002	0.002
48	0.028	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.002	0.003
49	0.026	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003
50	0.027	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004
51	0.026	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003	-0.003	0.003
52	0.027	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004	-0.002	0.004
53	0.026	0.004	-0.003	0.004	-0.003	0.004	-0.003	0.004	-0.003	0.004
54	0.028	0.005	-0.002	0.005	-0.002	0.005	-0.002	0.005	-0.003	0.005

Table 73: Simulation 1: Mean Bias (SE) in $\hat{\sigma}_D^2$ for Sample Size 34 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Complete Data Set										
1	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.002	0.000	0.001	0.000	0.002	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	-0.003	0.000	0.000	0.000
7	-0.001	0.001	-0.001	0.001	0.009	0.000	0.006	0.000	0.008	0.000
8	-0.001	0.001	-0.001	0.001	0.001	0.001	-0.003	0.001	0.001	0.001
9	-0.001	0.001	-0.001	0.001	0.012	0.000	0.008	0.000	0.010	0.000
10	-0.001	0.001	-0.001	0.001	0.000	0.001	-0.009	0.001	0.000	0.001
11	-0.001	0.001	-0.002	0.001	0.016	0.001	0.011	0.001	0.014	0.001
12	-0.001	0.002	-0.001	0.002	0.000	0.002	-0.020	0.002	-0.001	0.002
13	-0.003	0.002	-0.003	0.002	0.027	0.001	0.020	0.001	0.023	0.001
14	-0.002	0.003	-0.002	0.003	0.001	0.003	-0.013	0.002	0.000	0.003
15	-0.003	0.002	-0.003	0.002	0.032	0.001	0.023	0.001	0.028	0.001
16	-0.002	0.004	-0.002	0.004	0.000	0.004	-0.025	0.003	-0.001	0.004
17	-0.004	0.004	-0.004	0.003	0.049	0.002	0.034	0.002	0.041	0.002
18	-0.002	0.006	-0.002	0.006	-0.001	0.006	-0.064	0.005	-0.003	0.006
19	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	0.002	0.000	0.001	0.000	0.002	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	-0.003	0.000	0.000	0.000
25	-0.001	0.001	-0.001	0.001	0.009	0.000	0.006	0.000	0.008	0.000
26	-0.001	0.001	-0.001	0.001	0.001	0.001	-0.003	0.001	0.001	0.001
27	-0.001	0.001	-0.001	0.001	0.012	0.000	0.008	0.000	0.010	0.000
28	-0.001	0.001	-0.001	0.001	0.000	0.001	-0.009	0.001	0.000	0.001
29	-0.001	0.001	-0.002	0.001	0.016	0.001	0.011	0.001	0.014	0.001
30	-0.001	0.002	-0.001	0.002	0.000	0.002	-0.020	0.002	-0.001	0.002
31	-0.003	0.002	-0.003	0.002	0.027	0.001	0.020	0.001	0.023	0.001
32	-0.002	0.003	-0.002	0.003	0.001	0.003	-0.013	0.002	0.000	0.003
33	-0.003	0.002	-0.003	0.002	0.032	0.001	0.023	0.001	0.028	0.001
34	-0.002	0.004	-0.002	0.004	0.000	0.004	-0.025	0.003	-0.001	0.004
35	-0.004	0.004	-0.004	0.003	0.049	0.002	0.034	0.002	0.041	0.002
36	-0.002	0.006	-0.002	0.006	-0.001	0.006	-0.064	0.005	-0.003	0.006
37	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.002	0.000	0.001	0.000	0.002	0.000
42	0.000	0.000	0.000	0.000	0.000	0.000	-0.003	0.000	0.000	0.000
43	-0.001	0.001	-0.001	0.001	0.009	0.000	0.006	0.000	0.008	0.000
44	-0.001	0.001	-0.001	0.001	0.001	0.001	-0.003	0.001	0.001	0.001
45	-0.001	0.001	-0.001	0.001	0.012	0.000	0.008	0.000	0.010	0.000
46	-0.001	0.001	-0.001	0.001	0.000	0.001	-0.009	0.001	0.000	0.001
47	-0.001	0.001	-0.002	0.001	0.016	0.001	0.011	0.001	0.014	0.001
48	-0.001	0.002	-0.001	0.002	0.000	0.002	-0.020	0.002	-0.001	0.002
49	-0.003	0.002	-0.003	0.002	0.027	0.001	0.020	0.001	0.023	0.001
50	-0.002	0.003	-0.002	0.003	0.001	0.003	-0.013	0.002	0.000	0.003
51	-0.003	0.002	-0.003	0.002	0.032	0.001	0.023	0.001	0.028	0.001
52	-0.002	0.004	-0.002	0.004	0.000	0.004	-0.025	0.003	-0.001	0.004
53	-0.004	0.004	-0.004	0.003	0.049	0.002	0.034	0.002	0.041	0.002
54	-0.002	0.006	-0.002	0.006	-0.001	0.006	-0.064	0.005	-0.003	0.006

Table 73: Simulation 1: Mean Bias (SE) in $\hat{\sigma}_D^2$ for Sample Size 34 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Substantial Missing Data										
1	0.114	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
2	0.114	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
3	0.115	0.001	0.000	0.000	0.002	0.000	0.001	0.000	0.001	0.000
4	0.115	0.001	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
5	0.115	0.001	0.000	0.000	0.003	0.000	0.002	0.000	0.002	0.000
6	0.116	0.001	0.000	0.000	0.000	0.000	-0.004	0.000	0.000	0.000
7	0.124	0.002	0.000	0.001	0.011	0.000	0.008	0.000	0.010	0.000
8	0.125	0.002	0.000	0.001	0.003	0.001	-0.002	0.001	0.002	0.001
9	0.127	0.002	0.000	0.001	0.015	0.001	0.010	0.001	0.013	0.001
10	0.128	0.002	-0.001	0.002	0.001	0.002	-0.010	0.001	0.000	0.002
11	0.131	0.002	-0.001	0.001	0.020	0.001	0.014	0.001	0.017	0.001
12	0.132	0.003	-0.001	0.002	0.000	0.002	-0.023	0.002	-0.001	0.002
13	0.148	0.003	-0.001	0.002	0.035	0.001	0.025	0.001	0.030	0.001
14	0.149	0.004	-0.001	0.003	0.006	0.003	-0.013	0.003	0.004	0.003
15	0.153	0.004	-0.001	0.003	0.041	0.002	0.029	0.002	0.035	0.002
16	0.154	0.006	-0.002	0.004	0.006	0.005	-0.027	0.004	0.001	0.004
17	0.166	0.005	-0.002	0.004	0.060	0.002	0.040	0.002	0.050	0.002
18	0.167	0.008	-0.003	0.007	0.000	0.007	-0.073	0.006	-0.004	0.007
19	0.114	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
20	0.114	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
21	0.115	0.001	0.000	0.000	0.002	0.000	0.001	0.000	0.001	0.000
22	0.115	0.001	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
23	0.115	0.001	0.000	0.000	0.003	0.000	0.002	0.000	0.002	0.000
24	0.116	0.001	0.000	0.000	0.000	0.000	-0.004	0.000	0.000	0.000
25	0.124	0.002	0.000	0.001	0.011	0.000	0.008	0.000	0.010	0.000
26	0.125	0.002	0.000	0.001	0.003	0.001	-0.002	0.001	0.002	0.001
27	0.127	0.002	0.000	0.001	0.015	0.001	0.010	0.001	0.013	0.001
28	0.128	0.002	-0.001	0.002	0.001	0.002	-0.010	0.001	0.000	0.002
29	0.131	0.002	-0.001	0.001	0.020	0.001	0.014	0.001	0.017	0.001
30	0.132	0.003	-0.001	0.002	0.000	0.002	-0.023	0.002	-0.001	0.002
31	0.148	0.003	-0.001	0.002	0.035	0.001	0.025	0.001	0.030	0.001
32	0.149	0.004	-0.001	0.003	0.006	0.003	-0.013	0.003	0.004	0.003
33	0.153	0.004	-0.001	0.003	0.041	0.002	0.029	0.002	0.035	0.002
34	0.154	0.006	-0.002	0.004	0.006	0.005	-0.027	0.004	0.001	0.004
35	0.166	0.005	-0.002	0.004	0.060	0.002	0.040	0.002	0.050	0.002
36	0.167	0.008	-0.003	0.007	0.000	0.007	-0.073	0.006	-0.004	0.007
37	0.114	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
38	0.114	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
39	0.115	0.001	0.000	0.000	0.002	0.000	0.001	0.000	0.001	0.000
40	0.115	0.001	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
41	0.115	0.001	0.000	0.000	0.003	0.000	0.002	0.000	0.002	0.000
42	0.116	0.001	0.000	0.000	0.000	0.000	-0.004	0.000	0.000	0.000
43	0.124	0.002	0.000	0.001	0.011	0.000	0.008	0.000	0.010	0.000
44	0.125	0.002	0.000	0.001	0.003	0.001	-0.002	0.001	0.002	0.001
45	0.127	0.002	0.000	0.001	0.015	0.001	0.010	0.001	0.013	0.001
46	0.128	0.002	-0.001	0.002	0.001	0.002	-0.010	0.001	0.000	0.002
47	0.131	0.002	-0.001	0.001	0.020	0.001	0.014	0.001	0.017	0.001
48	0.132	0.003	-0.001	0.002	0.000	0.002	-0.023	0.002	-0.001	0.002
49	0.148	0.003	-0.001	0.002	0.035	0.001	0.025	0.001	0.030	0.001
50	0.149	0.004	-0.001	0.003	0.006	0.003	-0.013	0.003	0.004	0.003
51	0.153	0.004	-0.001	0.003	0.041	0.002	0.029	0.002	0.035	0.002
52	0.154	0.006	-0.002	0.004	0.006	0.005	-0.027	0.004	0.001	0.004
53	0.166	0.005	-0.002	0.004	0.060	0.002	0.040	0.002	0.050	0.002
54	0.167	0.008	-0.003	0.007	0.000	0.007	-0.073	0.006	-0.004	0.007

Table 74: Simulation 1: Mean Bias (SE) in $\hat{\sigma}^2_{WT}$ for Sample Size 34 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Complete Data Set										
1	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.004	0.001
8	0.000	0.001	0.000	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
9	0.000	0.001	0.000	0.001	-0.004	0.001	-0.004	0.001	-0.003	0.001
10	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
11	0.000	0.001	0.000	0.001	-0.003	0.001	-0.003	0.001	-0.003	0.001
12	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
13	0.001	0.002	0.001	0.002	-0.013	0.002	-0.013	0.002	-0.011	0.002
14	0.001	0.002	0.001	0.002	0.000	0.002	0.000	0.002	0.000	0.002
15	0.001	0.002	0.001	0.002	-0.011	0.002	-0.011	0.002	-0.010	0.002
16	0.001	0.002	0.001	0.002	0.000	0.002	0.000	0.002	0.001	0.002
17	0.001	0.002	0.001	0.002	-0.007	0.002	-0.007	0.002	-0.007	0.002
18	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002
19	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.004	0.001
26	0.000	0.001	0.000	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
27	0.000	0.001	0.000	0.001	-0.004	0.001	-0.004	0.001	-0.003	0.001
28	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
29	0.000	0.001	0.000	0.001	-0.003	0.001	-0.003	0.001	-0.003	0.001
30	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
31	0.001	0.002	0.001	0.002	-0.013	0.002	-0.013	0.002	-0.011	0.002
32	0.001	0.002	0.001	0.002	0.000	0.002	0.000	0.002	0.000	0.002
33	0.001	0.002	0.001	0.002	-0.011	0.002	-0.011	0.002	-0.010	0.002
34	0.001	0.002	0.001	0.002	0.000	0.002	0.000	0.002	0.001	0.002
35	0.001	0.002	0.001	0.002	-0.007	0.002	-0.007	0.002	-0.007	0.002
36	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002
37	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
42	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
43	0.000	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.004	0.001
44	0.000	0.001	0.000	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
45	0.000	0.001	0.000	0.001	-0.004	0.001	-0.004	0.001	-0.003	0.001
46	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
47	0.000	0.001	0.000	0.001	-0.003	0.001	-0.003	0.001	-0.003	0.001
48	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
49	0.001	0.002	0.001	0.002	-0.013	0.002	-0.013	0.002	-0.011	0.002
50	0.001	0.002	0.001	0.002	0.000	0.002	0.000	0.002	0.000	0.002
51	0.001	0.002	0.001	0.002	-0.011	0.002	-0.011	0.002	-0.010	0.002
52	0.001	0.002	0.001	0.002	0.000	0.002	0.000	0.002	0.001	0.002
53	0.001	0.002	0.001	0.002	-0.007	0.002	-0.007	0.002	-0.007	0.002
54	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002

Table 74: Simulation 1: Mean Bias (SE) in $\hat{\sigma}^2_{WT}$ for Sample Size 34 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Substantial Missing Data										
1	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
8	0.000	0.001	0.000	0.001	-0.002	0.001	-0.002	0.001	-0.001	0.001
9	0.000	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.004	0.001
10	0.000	0.001	0.000	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
11	0.000	0.001	0.000	0.001	-0.004	0.001	-0.004	0.001	-0.004	0.001
12	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
13	0.000	0.002	0.001	0.002	-0.016	0.002	-0.016	0.002	-0.014	0.002
14	0.000	0.002	0.000	0.002	-0.003	0.002	-0.003	0.002	-0.002	0.002
15	0.000	0.002	0.001	0.002	-0.014	0.002	-0.014	0.002	-0.012	0.002
16	0.000	0.002	0.000	0.002	-0.002	0.002	-0.001	0.002	-0.001	0.002
17	0.000	0.002	0.001	0.002	-0.010	0.002	-0.010	0.002	-0.010	0.002
18	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
19	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
26	0.000	0.001	0.000	0.001	-0.002	0.001	-0.002	0.001	-0.001	0.001
27	0.000	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.004	0.001
28	0.000	0.001	0.000	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
29	0.000	0.001	0.000	0.001	-0.004	0.001	-0.004	0.001	-0.004	0.001
30	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
31	0.000	0.002	0.001	0.002	-0.016	0.002	-0.016	0.002	-0.014	0.002
32	0.000	0.002	0.000	0.002	-0.003	0.002	-0.003	0.002	-0.002	0.002
33	0.000	0.002	0.001	0.002	-0.014	0.002	-0.014	0.002	-0.012	0.002
34	0.000	0.002	0.000	0.002	-0.002	0.002	-0.001	0.002	-0.001	0.002
35	0.000	0.002	0.001	0.002	-0.010	0.002	-0.010	0.002	-0.010	0.002
36	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
37	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
42	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
43	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
44	0.000	0.001	0.000	0.001	-0.002	0.001	-0.002	0.001	-0.001	0.001
45	0.000	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.004	0.001
46	0.000	0.001	0.000	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
47	0.000	0.001	0.000	0.001	-0.004	0.001	-0.004	0.001	-0.004	0.001
48	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
49	0.000	0.002	0.001	0.002	-0.016	0.002	-0.016	0.002	-0.014	0.002
50	0.000	0.002	0.000	0.002	-0.003	0.002	-0.003	0.002	-0.002	0.002
51	0.000	0.002	0.001	0.002	-0.014	0.002	-0.014	0.002	-0.012	0.002
52	0.000	0.002	0.000	0.002	-0.002	0.002	-0.001	0.002	-0.001	0.002
53	0.000	0.002	0.001	0.002	-0.010	0.002	-0.010	0.002	-0.010	0.002
54	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002

Table 75: Simulation 1: Mean Bias (SE) in $\hat{\sigma}^2_{WR}$ for Sample Size 34 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Complete Data Set										
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.001	0.000	0.001	-0.004	0.001	-0.004	0.001	-0.004	0.001
10	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
11	0.000	0.001	0.000	0.001	-0.009	0.001	-0.009	0.001	-0.007	0.001
12	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
13	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
14	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
15	0.000	0.002	0.000	0.002	-0.011	0.002	-0.011	0.002	-0.010	0.002
16	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
17	0.000	0.004	0.000	0.004	-0.029	0.003	-0.029	0.003	-0.023	0.003
18	0.000	0.004	0.000	0.004	-0.001	0.004	-0.001	0.004	0.001	0.004
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
26	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
27	0.000	0.001	0.000	0.001	-0.004	0.001	-0.004	0.001	-0.004	0.001
28	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
29	0.000	0.001	0.000	0.001	-0.009	0.001	-0.009	0.001	-0.007	0.001
30	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
31	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
32	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
33	0.000	0.002	0.000	0.002	-0.011	0.002	-0.011	0.002	-0.010	0.002
34	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
35	0.000	0.004	0.000	0.004	-0.029	0.003	-0.029	0.003	-0.023	0.003
36	0.000	0.004	0.000	0.004	-0.001	0.004	-0.001	0.004	0.001	0.004
37	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
42	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
43	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
44	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
45	0.000	0.001	0.000	0.001	-0.004	0.001	-0.004	0.001	-0.004	0.001
46	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
47	0.000	0.001	0.000	0.001	-0.009	0.001	-0.009	0.001	-0.007	0.001
48	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
49	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
50	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
51	0.000	0.002	0.000	0.002	-0.011	0.002	-0.011	0.002	-0.010	0.002
52	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
53	0.000	0.004	0.000	0.004	-0.029	0.003	-0.029	0.003	-0.023	0.003
54	0.000	0.004	0.000	0.004	-0.001	0.004	-0.001	0.004	0.001	0.004

Table 75: Simulation 1: Mean Bias (SE) in $\hat{\sigma}^2_{WR}$ for Sample Size 34 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Substantial Missing Data										
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.001	0.000	0.001	-0.004	0.001	-0.004	0.001	-0.004	0.001
10	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
11	0.000	0.001	0.000	0.001	-0.009	0.001	-0.009	0.001	-0.007	0.001
12	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
13	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
14	0.000	0.001	0.000	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
15	0.000	0.002	0.001	0.002	-0.012	0.002	-0.011	0.002	-0.010	0.002
16	0.000	0.002	0.001	0.002	-0.001	0.002	-0.001	0.002	0.000	0.002
17	0.000	0.004	0.001	0.004	-0.029	0.004	-0.029	0.004	-0.023	0.004
18	0.000	0.004	0.001	0.004	-0.001	0.004	-0.001	0.004	0.001	0.004
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
26	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
27	0.000	0.001	0.000	0.001	-0.004	0.001	-0.004	0.001	-0.004	0.001
28	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
29	0.000	0.001	0.000	0.001	-0.009	0.001	-0.009	0.001	-0.007	0.001
30	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
31	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
32	0.000	0.001	0.000	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
33	0.000	0.002	0.001	0.002	-0.012	0.002	-0.011	0.002	-0.010	0.002
34	0.000	0.002	0.001	0.002	-0.001	0.002	-0.001	0.002	0.000	0.002
35	0.000	0.004	0.001	0.004	-0.029	0.004	-0.029	0.004	-0.023	0.004
36	0.000	0.004	0.001	0.004	-0.001	0.004	-0.001	0.004	0.001	0.004
37	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
42	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
43	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
44	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
45	0.000	0.001	0.000	0.001	-0.004	0.001	-0.004	0.001	-0.004	0.001
46	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
47	0.000	0.001	0.000	0.001	-0.009	0.001	-0.009	0.001	-0.007	0.001
48	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
49	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
50	0.000	0.001	0.000	0.001	-0.001	0.001	-0.001	0.001	0.000	0.001
51	0.000	0.002	0.001	0.002	-0.012	0.002	-0.011	0.002	-0.010	0.002
52	0.000	0.002	0.001	0.002	-0.001	0.002	-0.001	0.002	0.000	0.002
53	0.000	0.004	0.001	0.004	-0.029	0.004	-0.029	0.004	-0.023	0.004
54	0.000	0.004	0.001	0.004	-0.001	0.004	-0.001	0.004	0.001	0.004

Table 76: Simulation 1: Percentage of ABE Failures for Sample Size 34 (1000 runs per simulation)

Sim	MoM	UN	CSH	FA0(2)	RIS
Complete Data Set					
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0.9	0.9	0.9	0
7	0.1	1.6	1.6	1.6	0.1
8	0.9	9.8	9.8	9.8	1.7
9	0.7	1.9	2.1	2.1	1
10	11.8	24.9	24.9	24.9	15
11	4.8	6.7	7.8	7.8	6.6
12	35.4	53.3	53.3	53.2	42.7
13	24.2	29.9	32.1	32.1	28
14	55.6	70.4	70.6	70.6	60.6
15	39	44.6	47.2	47.2	44.2
16	78.5	87.2	87.3	87.3	82.9
17	66.2	69.5	72.6	72.6	71.8
18	98.3	99.3	99.3	99.3	99.1
19	94.8	95.3	96.1	96.1	96.2
20	94.9	96.1	96.6	96.6	95.6
21	94.4	94.7	96.2	96.2	95.8
22	96.1	97.2	97.2	97.1	96.4
23	93.9	94.3	95.5	95.5	95.4
24	95.6	97.1	97.1	97.1	96.3
25	95.4	95.5	96.1	96.1	96.3
26	95.2	96.5	96.8	96.8	95.7
27	94.5	94.8	96.2	96.2	95.8
28	96.1	97.2	97.2	97.1	96.4
29	94.3	94.5	95.6	95.6	95.4
30	95.8	97.6	97.6	97.6	96.4
31	95.1	95.4	96.3	96.3	96.1
32	95.5	97.2	97.3	97.3	96.2
33	94.5	94.8	96.2	96.2	95.8
34	97.5	98.7	98.7	98.6	97.9
35	94.5	95.1	96.7	96.7	96.4
36	99.4	99.6	99.6	99.6	99.5
37	100	100	100	100	100
38	100	100	100	100	100
39	100	100	100	100	100
40	100	100	100	100	100
41	100	100	100	100	100
42	100	100	100	100	100
43	100	100	100	100	100
44	100	100	100	100	100
45	100	100	100	100	100
46	100	100	100	100	100
47	100	100	100	100	100
48	100	100	100	100	100
49	100	100	100	100	100
50	100	100	100	100	100
51	100	100	100	100	100
52	100	100	100	100	100
53	100	100	100	100	100
54	100	100	100	100	100
Substantial Missing Data					

Table 76: Simulation 1: Percentage of ABE Failures for Sample Size 34 (1000 runs per simulation)

Sim	MoM	UN	CSH	FA0(2)	RIS
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0.1	0.6	0.6	0.6	0
5	0	0.2	0.2	0.2	0
6	1.7	4.4	4.4	4.4	0.4
7	12.1	6	6.2	5.8	1.3
8	20.1	21.8	21.7	21.4	8.3
9	20.5	10.1	10.5	10.4	5.9
10	43.6	43.4	43.4	43.3	31.2
11	35.3	19.7	21.8	21.5	18.1
12	66.5	70.6	70.6	70.6	62.6
13	60	51.6	55.7	55.3	53.5
14	81.7	85.5	86.1	85.5	82.4
15	71.8	63.8	68.5	67.6	67.8
16	92.3	93.6	93.9	93.7	93
17	89	84.4	89.4	89.1	91.4
18	98.9	99.4	99.6	99.6	99.5
19	100	95.2	96.6	96.5	96.7
20	100	96.2	96.7	96.6	96
21	100	94.8	95.9	95.9	95.6
22	100	96.5	96.7	96.7	96.3
23	100	94.7	95.9	95.9	96
24	100	97.1	97.1	97	96.6
25	99.7	95	96.5	96.4	96.2
26	99.1	96.2	96.6	96.6	96
27	99.5	94.8	95.9	95.9	95.6
28	98.7	96.5	96.7	96.7	96.3
29	98.5	95	96.1	96.1	96.3
30	97.9	97.1	97.3	97.2	96.9
31	97.8	95	96.2	96.1	96.1
32	97.8	97.4	97.8	97.6	97.6
33	97.5	95.1	96.5	96.5	96.2
34	98.5	98.9	99.1	99.1	98.7
35	98.5	97.2	98.4	98.4	98.6
36	99.9	99.9	99.9	99.9	99.9
37	100	100	100	100	100
38	100	100	100	100	100
39	100	100	100	100	100
40	100	100	100	100	100
41	100	100	100	100	100
42	100	100	100	100	100
43	100	100	100	100	100
44	100	100	100	100	100
45	100	100	100	100	100
46	100	100	100	100	100
47	100	100	100	100	100
48	100	100	100	100	100
49	100	100	100	100	100
50	100	100	100	100	100
51	100	100	100	100	100
52	100	100	100	100	100
53	100	100	100	100	100
54	100	100	100	100	100

Table 77: Simulation 1: Mean Bias (SE) in $\hat{\delta}$ for Sample Size 80 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Complete Data Set										
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
7	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
8	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
9	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
10	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
11	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
12	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
13	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
14	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
15	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
16	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
17	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
18	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
24	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
25	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
26	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
27	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
28	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
29	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
30	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
31	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
32	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
33	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
34	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
35	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
36	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003
37	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
42	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
43	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
44	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
45	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
46	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
47	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
48	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
49	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
50	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
51	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
52	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
53	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
54	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003

Table 77: Simulation 1: Mean Bias (SE) in $\hat{\delta}$ for Sample Size 80 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Substantial Missing Data										
1	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.012	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
7	0.012	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
8	0.012	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
9	0.012	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
10	0.012	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
11	0.012	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
12	0.011	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
13	0.012	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
14	0.011	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
15	0.012	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
16	0.011	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
17	0.011	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
18	0.010	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003
19	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
22	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
24	0.012	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
25	0.012	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
26	0.012	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
27	0.012	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
28	0.012	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
29	0.012	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
30	0.011	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
31	0.012	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
32	0.011	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
33	0.012	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
34	0.011	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
35	0.011	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
36	0.010	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003
37	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
38	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
40	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
41	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
42	0.012	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
43	0.012	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
44	0.012	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
45	0.012	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
46	0.012	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
47	0.012	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
48	0.011	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
49	0.012	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
50	0.011	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
51	0.012	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
52	0.011	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.001	0.002
53	0.011	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002
54	0.010	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003	-0.001	0.003

Table 78: Simulation 1: Mean Bias (SE) in $\hat{\sigma}_D^2$ for Sample Size 80 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Complete Data Set										
1	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	-0.002	0.000	0.000	0.000
7	-0.001	0.000	-0.001	0.000	0.005	0.000	0.005	0.000	0.005	0.000
8	0.000	0.001	0.000	0.001	0.000	0.001	-0.002	0.001	0.000	0.001
9	-0.001	0.001	-0.001	0.001	0.007	0.000	0.006	0.000	0.006	0.000
10	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.006	0.001	-0.001	0.001
11	-0.001	0.001	-0.001	0.001	0.010	0.000	0.008	0.000	0.009	0.000
12	-0.001	0.001	0.000	0.001	0.000	0.001	-0.014	0.001	-0.001	0.001
13	-0.002	0.001	-0.002	0.001	0.017	0.001	0.014	0.001	0.015	0.001
14	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.009	0.002	-0.001	0.002
15	-0.002	0.002	-0.002	0.002	0.020	0.001	0.016	0.001	0.018	0.001
16	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.016	0.002	-0.002	0.002
17	-0.003	0.002	-0.003	0.002	0.030	0.001	0.024	0.001	0.027	0.001
18	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.047	0.003	-0.002	0.004
19	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	-0.002	0.000	0.000	0.000
25	-0.001	0.000	-0.001	0.000	0.005	0.000	0.005	0.000	0.005	0.000
26	0.000	0.001	0.000	0.001	0.000	0.001	-0.002	0.001	0.000	0.001
27	-0.001	0.001	-0.001	0.001	0.007	0.000	0.006	0.000	0.006	0.000
28	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.006	0.001	-0.001	0.001
29	-0.001	0.001	-0.001	0.001	0.010	0.000	0.008	0.000	0.009	0.000
30	-0.001	0.001	0.000	0.001	0.000	0.001	-0.014	0.001	-0.001	0.001
31	-0.002	0.001	-0.002	0.001	0.017	0.001	0.014	0.001	0.015	0.001
32	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.009	0.002	-0.001	0.002
33	-0.002	0.002	-0.002	0.002	0.020	0.001	0.016	0.001	0.018	0.001
34	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.016	0.002	-0.002	0.002
35	-0.003	0.002	-0.003	0.002	0.030	0.001	0.024	0.001	0.027	0.001
36	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.047	0.003	-0.002	0.004
37	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
42	0.000	0.000	0.000	0.000	0.000	0.000	-0.002	0.000	0.000	0.000
43	-0.001	0.000	-0.001	0.000	0.005	0.000	0.005	0.000	0.005	0.000
44	0.000	0.001	0.000	0.001	0.000	0.001	-0.002	0.001	0.000	0.001
45	-0.001	0.001	-0.001	0.001	0.007	0.000	0.006	0.000	0.006	0.000
46	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.006	0.001	-0.001	0.001
47	-0.001	0.001	-0.001	0.001	0.010	0.000	0.008	0.000	0.009	0.000
48	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.014	0.001	-0.001	0.001
49	-0.002	0.001	-0.002	0.001	0.017	0.001	0.014	0.001	0.015	0.001
50	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.009	0.002	-0.001	0.002
51	-0.002	0.002	-0.002	0.002	0.020	0.001	0.016	0.001	0.018	0.001
52	-0.001	0.002	-0.001	0.002	-0.001	0.002	-0.016	0.002	-0.002	0.002
53	-0.003	0.002	-0.003	0.002	0.030	0.001	0.024	0.001	0.027	0.001
54	-0.001	0.004	-0.001	0.004	-0.001	0.004	-0.047	0.003	-0.002	0.004

Table 78: Simulation 1: Mean Bias (SE) in $\hat{\sigma}_D^2$ for Sample Size 80 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Substantial Missing Data										
1	0.041	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000
2	0.041	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.042	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
4	0.042	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
5	0.042	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
6	0.042	0.000	0.000	0.000	0.000	0.000	-0.002	0.000	0.000	0.000
7	0.044	0.001	0.000	0.000	0.006	0.000	0.005	0.000	0.005	0.000
8	0.045	0.001	0.000	0.001	0.000	0.001	-0.002	0.001	0.000	0.001
9	0.045	0.001	0.000	0.001	0.008	0.000	0.006	0.000	0.007	0.000
10	0.046	0.001	0.000	0.001	0.000	0.001	-0.006	0.001	0.000	0.001
11	0.047	0.001	-0.001	0.001	0.011	0.000	0.009	0.000	0.010	0.000
12	0.047	0.002	0.000	0.001	0.000	0.001	-0.014	0.001	0.000	0.001
13	0.052	0.002	-0.001	0.001	0.018	0.001	0.014	0.001	0.017	0.001
14	0.052	0.002	0.000	0.002	0.000	0.002	-0.009	0.002	0.000	0.002
15	0.054	0.002	-0.001	0.002	0.022	0.001	0.017	0.001	0.020	0.001
16	0.054	0.003	0.000	0.003	0.000	0.003	-0.016	0.002	0.000	0.003
17	0.058	0.003	-0.002	0.002	0.033	0.001	0.026	0.001	0.030	0.001
18	0.060	0.004	0.001	0.004	0.001	0.004	-0.046	0.004	0.001	0.004
19	0.041	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000
20	0.041	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.042	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
22	0.042	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
23	0.042	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
24	0.042	0.000	0.000	0.000	0.000	0.000	-0.002	0.000	0.000	0.000
25	0.044	0.001	0.000	0.000	0.006	0.000	0.005	0.000	0.005	0.000
26	0.045	0.001	0.000	0.001	0.000	0.001	-0.002	0.001	0.000	0.001
27	0.045	0.001	0.000	0.001	0.008	0.000	0.006	0.000	0.007	0.000
28	0.046	0.001	0.000	0.001	0.000	0.001	-0.006	0.001	0.000	0.001
29	0.047	0.001	-0.001	0.001	0.011	0.000	0.009	0.000	0.010	0.000
30	0.047	0.002	0.000	0.001	0.000	0.001	-0.014	0.001	0.000	0.001
31	0.052	0.002	-0.001	0.001	0.018	0.001	0.014	0.001	0.017	0.001
32	0.052	0.002	0.000	0.002	0.000	0.002	-0.009	0.002	0.000	0.002
33	0.054	0.002	-0.001	0.002	0.022	0.001	0.017	0.001	0.020	0.001
34	0.054	0.003	0.000	0.003	0.000	0.003	-0.016	0.002	0.000	0.003
35	0.058	0.003	-0.002	0.002	0.033	0.001	0.026	0.001	0.030	0.001
36	0.060	0.004	0.001	0.004	0.001	0.004	-0.046	0.004	0.001	0.004
37	0.041	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000
38	0.041	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.042	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
40	0.042	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.000
41	0.042	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000
42	0.042	0.000	0.000	0.000	0.000	0.000	-0.002	0.000	0.000	0.000
43	0.044	0.001	0.000	0.000	0.006	0.000	0.005	0.000	0.005	0.000
44	0.045	0.001	0.000	0.001	0.000	0.001	-0.002	0.001	0.000	0.001
45	0.045	0.001	0.000	0.001	0.008	0.000	0.006	0.000	0.007	0.000
46	0.046	0.001	0.000	0.001	0.000	0.001	-0.006	0.001	0.000	0.001
47	0.047	0.001	-0.001	0.001	0.011	0.000	0.009	0.000	0.010	0.000
48	0.047	0.002	0.000	0.001	0.000	0.001	-0.014	0.001	0.000	0.001
49	0.052	0.002	-0.001	0.001	0.018	0.001	0.014	0.001	0.017	0.001
50	0.052	0.002	0.000	0.002	0.000	0.002	-0.009	0.002	0.000	0.002
51	0.054	0.002	-0.001	0.002	0.022	0.001	0.017	0.001	0.020	0.001
52	0.054	0.003	0.000	0.003	0.000	0.003	-0.016	0.002	0.000	0.003
53	0.058	0.003	-0.002	0.002	0.033	0.001	0.026	0.001	0.030	0.001
54	0.060	0.004	0.001	0.004	0.001	0.004	-0.046	0.004	0.001	0.004

Table 79: Simulation 1: Mean Bias (SE) in $\hat{\sigma}^2_{WT}$ for Sample Size 80 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Complete Data Set										
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	-0.003	0.000	-0.003	0.000	-0.003	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	-0.003	0.000	-0.003	0.000	-0.002	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
13	0.000	0.001	0.000	0.001	-0.009	0.001	-0.009	0.001	-0.008	0.001
14	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
15	0.000	0.001	0.000	0.001	-0.007	0.001	-0.007	0.001	-0.007	0.001
16	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
17	0.000	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.005	0.001
18	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.000	0.000	0.000	-0.003	0.000	-0.003	0.000	-0.003	0.000
26	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
27	0.000	0.000	0.000	0.000	-0.003	0.000	-0.003	0.000	-0.002	0.000
28	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
29	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
30	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
31	0.000	0.001	0.000	0.001	-0.009	0.001	-0.009	0.001	-0.008	0.001
32	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
33	0.000	0.001	0.000	0.001	-0.007	0.001	-0.007	0.001	-0.007	0.001
34	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
35	0.000	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.005	0.001
36	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
37	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
42	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
43	0.000	0.000	0.000	0.000	-0.003	0.000	-0.003	0.000	-0.003	0.000
44	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
45	0.000	0.000	0.000	0.000	-0.003	0.000	-0.003	0.000	-0.002	0.000
46	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
47	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
48	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
49	0.000	0.001	0.000	0.001	-0.009	0.001	-0.009	0.001	-0.008	0.001
50	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
51	0.000	0.001	0.000	0.001	-0.007	0.001	-0.007	0.001	-0.007	0.001
52	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
53	0.000	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.005	0.001
54	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001

Table 79: Simulation 1: Mean Bias (SE) in $\hat{\sigma}^2_{WT}$ for Sample Size 80 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Substantial Missing Data										
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	-0.004	0.000	-0.004	0.000	-0.003	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	-0.003	0.000	-0.003	0.000	-0.003	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
13	0.000	0.001	0.000	0.001	-0.009	0.001	-0.009	0.001	-0.009	0.001
14	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
15	0.000	0.001	0.000	0.001	-0.008	0.001	-0.008	0.001	-0.007	0.001
16	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
17	0.000	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.006	0.001
18	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.000	0.000	0.000	-0.004	0.000	-0.004	0.000	-0.003	0.000
26	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
27	0.000	0.000	0.000	0.000	-0.003	0.000	-0.003	0.000	-0.003	0.000
28	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
29	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
30	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
31	0.000	0.001	0.000	0.001	-0.009	0.001	-0.009	0.001	-0.009	0.001
32	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
33	0.000	0.001	0.000	0.001	-0.008	0.001	-0.008	0.001	-0.007	0.001
34	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
35	0.000	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.006	0.001
36	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
37	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
42	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
43	0.000	0.000	0.000	0.000	-0.004	0.000	-0.004	0.000	-0.003	0.000
44	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
45	0.000	0.000	0.000	0.000	-0.003	0.000	-0.003	0.000	-0.003	0.000
46	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
47	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
48	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
49	0.000	0.001	0.000	0.001	-0.009	0.001	-0.009	0.001	-0.009	0.001
50	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
51	0.000	0.001	0.000	0.001	-0.008	0.001	-0.008	0.001	-0.007	0.001
52	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
53	0.000	0.001	0.000	0.001	-0.005	0.001	-0.005	0.001	-0.006	0.001
54	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001

Table 80: Simulation 1: Mean Bias (SE) in $\hat{\sigma}^2_{WR}$ for Sample Size 80 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Complete Data Set										
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
12	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
13	0.000	0.001	0.000	0.001	-0.003	0.001	-0.003	0.001	-0.003	0.001
14	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
15	0.001	0.001	0.001	0.001	-0.007	0.001	-0.007	0.001	-0.006	0.001
16	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
17	0.001	0.003	0.001	0.003	-0.017	0.002	-0.017	0.002	-0.014	0.002
18	0.001	0.003	0.001	0.003	0.001	0.003	0.001	0.003	0.001	0.002
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
26	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
27	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
28	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
29	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
30	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
31	0.000	0.001	0.000	0.001	-0.003	0.001	-0.003	0.001	-0.003	0.001
32	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
33	0.001	0.001	0.001	0.001	-0.007	0.001	-0.007	0.001	-0.006	0.001
34	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
35	0.001	0.003	0.001	0.003	-0.017	0.002	-0.017	0.002	-0.014	0.002
36	0.001	0.003	0.001	0.003	0.001	0.003	0.001	0.003	0.001	0.002
37	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
42	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
43	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
44	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
45	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
46	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
47	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
48	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
49	0.000	0.001	0.000	0.001	-0.003	0.001	-0.003	0.001	-0.003	0.001
50	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
51	0.001	0.001	0.001	0.001	-0.007	0.001	-0.007	0.001	-0.006	0.001
52	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
53	0.001	0.003	0.001	0.003	-0.017	0.002	-0.017	0.002	-0.014	0.002
54	0.001	0.003	0.001	0.003	0.001	0.003	0.001	0.003	0.001	0.002

Table 80: Simulation 1: Mean Bias (SE) in $\hat{\sigma}^2_{WR}$ for Sample Size 80 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE	FA0(2) Bias	FA0(2) SE	RIS Bias	RIS SE
Substantial Missing Data										
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
12	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
13	0.000	0.001	0.000	0.001	-0.003	0.001	-0.003	0.001	-0.003	0.001
14	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
15	0.001	0.001	0.001	0.001	-0.007	0.001	-0.007	0.001	-0.006	0.001
16	0.001	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
17	0.001	0.003	0.001	0.003	-0.018	0.002	-0.018	0.002	-0.015	0.002
18	0.001	0.003	0.001	0.003	0.001	0.003	0.001	0.003	0.001	0.003
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
26	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
27	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
28	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
29	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
30	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
31	0.000	0.001	0.000	0.001	-0.003	0.001	-0.003	0.001	-0.003	0.001
32	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
33	0.001	0.001	0.001	0.001	-0.007	0.001	-0.007	0.001	-0.006	0.001
34	0.001	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
35	0.001	0.003	0.001	0.003	-0.018	0.002	-0.018	0.002	-0.015	0.002
36	0.001	0.003	0.001	0.003	0.001	0.003	0.001	0.003	0.001	0.003
37	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
42	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
43	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
44	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
45	0.000	0.000	0.000	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000
46	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
47	0.000	0.001	0.000	0.001	-0.006	0.001	-0.006	0.001	-0.005	0.001
48	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
49	0.000	0.001	0.000	0.001	-0.003	0.001	-0.003	0.001	-0.003	0.001
50	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
51	0.001	0.001	0.001	0.001	-0.007	0.001	-0.007	0.001	-0.006	0.001
52	0.001	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
53	0.001	0.003	0.001	0.003	-0.018	0.002	-0.018	0.002	-0.015	0.002
54	0.001	0.003	0.001	0.003	0.001	0.003	0.001	0.003	0.001	0.003

Table 81: Simulation 1: Percentage of ABE Failures for Sample Size 80 (1000 runs per simulation)

Sim	MoM	UN	CSH	FA0(2)	RIS
Complete Data Set					
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	0	0	0
8	0	0	0	0	0
9	0	0	0	0	0
10	0.1	0.1	0.1	0.1	0.1
11	0	0	0	0	0
12	1.9	2.3	2.3	2.3	2.1
13	0.4	0.6	0.8	0.8	0.7
14	6.7	7.5	7.5	7.5	7.2
15	2.8	2.8	2.8	2.8	3
16	17.7	20.8	20.8	20.8	19.4
17	10.3	10.5	11.3	11.3	11.3
18	51.3	55.5	55.5	55.5	54.1
19	95.3	95.6	95.9	95.9	95.8
20	95.1	95.8	95.8	95.8	95
21	95.2	95.3	96	96	95.9
22	95.8	96.5	96.5	96.5	96.4
23	95.2	95.3	95.4	95.4	95.4
24	95.5	96.2	96.2	96.2	96.3
25	95.5	95.7	96.2	96.2	96.1
26	95.7	96.5	96.5	96.5	96.2
27	95.4	95.5	96	96	95.9
28	95.8	96.5	96.5	96.5	96.4
29	95.5	95.6	95.8	95.8	95.7
30	95.5	96.3	96.3	96.3	96
31	95.3	95.6	96.1	96.1	96.1
32	95.9	96.4	96.4	96.4	96
33	95.4	95.5	96	96	95.9
34	95.8	96.5	96.5	96.5	96.4
35	95.6	95.6	95.7	95.7	95.6
36	95.6	96.3	96.3	96.3	96.2
37	100	100	100	100	100
38	100	100	100	100	100
39	100	100	100	100	100
40	100	100	100	100	100
41	100	100	100	100	100
42	100	100	100	100	100
43	100	100	100	100	100
44	100	100	100	100	100
45	100	100	100	100	100
46	100	100	100	100	100
47	100	100	100	100	100
48	100	100	100	100	100
49	100	100	100	100	100
50	100	100	100	100	100
51	100	100	100	100	100
52	100	100	100	100	100
53	100	100	100	100	100
54	100	100	100	100	100
Substantial Missing Data					

Table 81: Simulation 1: Percentage of ABE Failures for Sample Size 80 (1000 runs per simulation)

Sim	MoM	UN	CSH	FA0(2)	RIS
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	0	0	0
8	0	0	0	0	0
9	0	0	0	0	0
10	0.3	0.3	0.2	0.2	0.1
11	0	0	0	0	0
12	4.2	4	4	4	3.8
13	2.6	1.3	1.4	1.4	1.3
14	12.4	11.2	11.2	11.2	10.6
15	5.4	4.1	4.5	4.4	4.4
16	25.3	26.3	26.3	26.3	24.8
17	17.9	15.8	16.8	16.7	17.2
18	58.6	63.2	63.2	63.2	60.9
19	100	95.7	96	96	95.8
20	100	95.7	95.7	95.7	95.2
21	100	95.2	95.8	95.7	95.6
22	100	95.7	95.7	95.7	95.4
23	100	95.6	96.1	96.1	96.1
24	99.9	96.5	96.5	96.5	96.3
25	99.1	95.6	96.3	96.2	96
26	98.7	95.8	95.8	95.8	95.7
27	98.6	95.3	95.8	95.7	95.6
28	98.1	95.7	95.7	95.7	95.4
29	98.1	95.5	95.8	95.8	96
30	97.7	96.4	96.4	96.4	96.3
31	97	95.2	96.1	96.1	95.9
32	96.9	95.5	95.5	95.5	95.3
33	96.9	95.3	95.8	95.7	95.6
34	97	95.7	95.7	95.7	95.4
35	96.7	95.6	95.8	95.9	96.2
36	96.5	96.9	96.9	96.9	96.8
37	100	100	100	100	100
38	100	100	100	100	100
39	100	100	100	100	100
40	100	100	100	100	100
41	100	100	100	100	100
42	100	100	100	100	100
43	100	100	100	100	100
44	100	100	100	100	100
45	100	100	100	100	100
46	100	100	100	100	100
47	100	100	100	100	100
48	100	100	100	100	100
49	100	100	100	100	100
50	100	100	100	100	100
51	100	100	100	100	100
52	100	100	100	100	100
53	100	100	100	100	100
54	100	100	100	100	100

Table 82: Simulation 1: Mean Bias (SE) in Estimated IBE FDA Metric for Sample Size 16 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	RIS Bias	RIS SE
Complete Data Set						
1	0.000	0.000	0.000	0.000	0.001	0.000
2	0.000	0.000	0.000	0.000	0.001	0.000
3	0.001	0.000	0.001	0.000	0.002	0.000
4	0.001	0.000	0.001	0.000	0.001	0.000
5	0.001	0.001	0.001	0.000	0.005	0.000
6	0.002	0.001	0.002	0.001	0.002	0.001
7	-0.010	0.002	-0.011	0.002	-0.001	0.001
8	-0.008	0.002	-0.008	0.002	-0.005	0.002
9	0.008	0.004	0.006	0.004	0.037	0.004
10	0.013	0.005	0.012	0.005	0.016	0.004
11	0.014	0.008	0.011	0.008	0.069	0.006
12	0.023	0.008	0.021	0.008	0.022	0.008
13	0.017	0.008	0.015	0.008	0.064	0.007
14	0.026	0.008	0.024	0.009	0.035	0.008
15	0.024	0.012	0.019	0.012	0.104	0.010
16	0.039	0.013	0.034	0.013	0.045	0.012
17	0.042	0.024	0.034	0.023	0.212	0.018
18	0.071	0.024	0.063	0.024	0.062	0.024
19	0.000	0.000	0.000	0.000	0.001	0.000
20	0.000	0.000	0.000	0.000	0.001	0.000
21	0.000	0.000	0.000	0.000	0.002	0.000
22	0.001	0.001	0.001	0.001	0.001	0.001
23	0.001	0.001	0.000	0.001	0.005	0.001
24	0.002	0.001	0.002	0.001	0.002	0.001
25	-0.011	0.002	-0.011	0.002	-0.001	0.002
26	-0.008	0.002	-0.008	0.002	-0.005	0.002
27	0.007	0.005	0.006	0.005	0.036	0.004
28	0.013	0.005	0.011	0.005	0.015	0.005
29	0.013	0.008	0.010	0.008	0.069	0.006
30	0.023	0.008	0.020	0.008	0.021	0.008
31	0.016	0.008	0.014	0.008	0.063	0.007
32	0.026	0.009	0.024	0.009	0.035	0.008
33	0.023	0.012	0.018	0.012	0.103	0.010
34	0.038	0.013	0.034	0.013	0.045	0.013
35	0.041	0.024	0.033	0.023	0.211	0.018
36	0.071	0.025	0.062	0.025	0.061	0.024
37	0.000	0.001	0.000	0.001	0.000	0.001
38	0.000	0.001	0.000	0.001	0.001	0.001
39	0.000	0.001	0.000	0.001	0.002	0.001
40	0.001	0.001	0.001	0.001	0.001	0.001
41	0.000	0.002	0.000	0.002	0.004	0.001
42	0.002	0.002	0.001	0.002	0.001	0.002
43	-0.012	0.003	-0.012	0.003	-0.002	0.003
44	-0.008	0.004	-0.009	0.004	-0.005	0.004
45	0.006	0.006	0.005	0.006	0.035	0.005
46	0.013	0.006	0.011	0.006	0.015	0.006
47	0.012	0.009	0.009	0.009	0.067	0.007
48	0.022	0.010	0.020	0.010	0.021	0.010
49	0.015	0.010	0.012	0.010	0.061	0.009
50	0.026	0.011	0.024	0.011	0.035	0.010
51	0.021	0.014	0.016	0.014	0.101	0.012
52	0.038	0.015	0.033	0.015	0.044	0.015
53	0.039	0.025	0.031	0.025	0.209	0.019
54	0.070	0.026	0.061	0.026	0.060	0.026

Table 82: Simulation 1: Mean Bias (SE) in Estimated IBE FDA Metric for Sample Size 16 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	RIS Bias	RIS SE
Substantial Missing Data						
1	0.368	0.001	0.002	0.000	0.002	0.000
2	0.369	0.002	0.002	0.000	0.002	0.000
3	0.370	0.002	0.003	0.000	0.005	0.000
4	0.372	0.002	0.004	0.000	0.004	0.000
5	0.372	0.002	0.003	0.001	0.009	0.001
6	0.375	0.003	0.005	0.001	0.005	0.001
7	0.393	0.005	-0.001	0.003	0.012	0.002
8	0.399	0.006	0.004	0.003	0.011	0.003
9	0.425	0.007	0.025	0.006	0.060	0.005
10	0.436	0.008	0.033	0.007	0.046	0.006
11	0.447	0.011	0.040	0.010	0.102	0.008
12	0.467	0.013	0.049	0.011	0.062	0.010
13	0.514	0.013	0.058	0.012	0.121	0.009
14	0.534	0.015	0.075	0.014	0.106	0.011
15	0.540	0.018	0.075	0.017	0.174	0.014
16	0.571	0.020	0.097	0.019	0.133	0.017
17	0.610	0.032	0.118	0.031	0.312	0.024
18	0.668	0.035	0.149	0.034	0.174	0.032
19	0.396	0.002	0.001	0.001	0.002	0.000
20	0.397	0.002	0.002	0.001	0.002	0.001
21	0.398	0.002	0.002	0.001	0.004	0.001
22	0.399	0.002	0.003	0.001	0.004	0.001
23	0.400	0.002	0.003	0.001	0.008	0.001
24	0.402	0.003	0.004	0.001	0.004	0.001
25	0.421	0.005	-0.002	0.003	0.011	0.003
26	0.426	0.006	0.002	0.004	0.009	0.003
27	0.452	0.008	0.023	0.006	0.058	0.005
28	0.463	0.009	0.030	0.007	0.043	0.006
29	0.475	0.012	0.038	0.010	0.100	0.008
30	0.492	0.013	0.047	0.012	0.059	0.011
31	0.541	0.014	0.057	0.012	0.118	0.010
32	0.560	0.015	0.072	0.014	0.102	0.012
33	0.567	0.019	0.072	0.017	0.171	0.014
34	0.596	0.021	0.093	0.020	0.129	0.017
35	0.636	0.033	0.115	0.031	0.308	0.024
36	0.692	0.036	0.145	0.035	0.169	0.032
37	0.455	0.002	0.001	0.002	0.001	0.001
38	0.455	0.002	0.000	0.002	0.001	0.002
39	0.456	0.003	0.001	0.002	0.003	0.002
40	0.457	0.003	0.001	0.002	0.002	0.002
41	0.458	0.003	0.002	0.002	0.007	0.002
42	0.459	0.004	0.002	0.003	0.002	0.003
43	0.479	0.007	-0.004	0.006	0.008	0.005
44	0.482	0.008	-0.003	0.007	0.005	0.006
45	0.510	0.010	0.020	0.008	0.054	0.007
46	0.518	0.011	0.026	0.010	0.037	0.009
47	0.532	0.014	0.035	0.012	0.096	0.010
48	0.546	0.016	0.042	0.014	0.053	0.013
49	0.598	0.017	0.052	0.015	0.112	0.013
50	0.614	0.019	0.064	0.017	0.094	0.015
51	0.624	0.021	0.067	0.020	0.164	0.016
52	0.649	0.024	0.085	0.023	0.119	0.020
53	0.692	0.035	0.110	0.033	0.302	0.027
54	0.742	0.039	0.136	0.038	0.159	0.035

Table 83: Simulation 1: Mean Bias (SE) in Estimated PBE FDA Metric for Sample Size 16 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE
Complete Data Set						
1	-0.008	0.001	-0.008	0.001	-0.008	0.001
2	-0.008	0.001	-0.008	0.001	-0.008	0.001
3	-0.003	0.001	-0.003	0.001	-0.003	0.001
4	-0.004	0.001	-0.004	0.001	-0.004	0.001
5	0.001	0.001	0.001	0.001	0.000	0.001
6	0.000	0.002	0.000	0.002	0.000	0.002
7	0.007	0.009	0.007	0.009	0.007	0.009
8	0.003	0.009	0.003	0.009	0.003	0.009
9	0.011	0.010	0.011	0.010	0.010	0.010
10	0.006	0.011	0.005	0.011	0.005	0.011
11	0.017	0.012	0.016	0.012	0.014	0.012
12	0.009	0.013	0.008	0.013	0.008	0.013
13	0.024	0.026	0.024	0.026	0.024	0.026
14	0.011	0.027	0.011	0.027	0.011	0.027
15	0.032	0.028	0.030	0.028	0.029	0.028
16	0.015	0.031	0.014	0.031	0.014	0.031
17	0.050	0.034	0.047	0.034	0.043	0.034
18	0.027	0.038	0.024	0.038	0.022	0.038
19	-0.008	0.001	-0.008	0.001	-0.008	0.001
20	-0.008	0.001	-0.008	0.001	-0.008	0.001
21	-0.003	0.001	-0.003	0.001	-0.003	0.001
22	-0.004	0.001	-0.004	0.001	-0.004	0.001
23	0.001	0.001	0.000	0.001	0.000	0.001
24	0.000	0.002	0.000	0.002	0.000	0.002
25	0.007	0.009	0.007	0.009	0.007	0.009
26	0.003	0.009	0.003	0.009	0.003	0.009
27	0.011	0.010	0.010	0.010	0.010	0.010
28	0.005	0.011	0.005	0.011	0.005	0.011
29	0.016	0.012	0.015	0.012	0.014	0.012
30	0.009	0.013	0.008	0.013	0.007	0.013
31	0.024	0.026	0.023	0.026	0.023	0.026
32	0.011	0.028	0.011	0.028	0.010	0.028
33	0.031	0.028	0.029	0.028	0.028	0.028
34	0.015	0.031	0.014	0.031	0.013	0.031
35	0.049	0.034	0.046	0.034	0.041	0.034
36	0.027	0.038	0.024	0.038	0.021	0.038
37	-0.008	0.001	-0.008	0.001	-0.008	0.001
38	-0.008	0.001	-0.008	0.001	-0.008	0.001
39	-0.004	0.001	-0.004	0.001	-0.004	0.001
40	-0.004	0.002	-0.004	0.002	-0.004	0.002
41	0.000	0.002	0.000	0.002	0.000	0.002
42	-0.001	0.003	-0.001	0.003	-0.001	0.003
43	0.006	0.009	0.006	0.009	0.006	0.009
44	0.003	0.010	0.003	0.010	0.003	0.010
45	0.010	0.011	0.009	0.011	0.009	0.011
46	0.005	0.012	0.004	0.012	0.004	0.012
47	0.015	0.012	0.014	0.012	0.012	0.012
48	0.008	0.014	0.007	0.015	0.007	0.015
49	0.022	0.026	0.022	0.026	0.021	0.026
50	0.011	0.028	0.010	0.028	0.010	0.028
51	0.029	0.028	0.027	0.028	0.026	0.028
52	0.015	0.032	0.013	0.032	0.013	0.032
53	0.046	0.034	0.043	0.034	0.038	0.034
54	0.026	0.040	0.022	0.040	0.020	0.040

Table 83: Simulation 1: Mean Bias (SE) in Estimated PBE FDA Metric for Sample Size 16 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE
Substantial Missing Data						
1	-0.212	0.004	-0.008	0.001	-0.007	0.001
2	-0.212	0.004	-0.008	0.001	-0.007	0.001
3	-0.214	0.005	-0.004	0.001	-0.002	0.001
4	-0.215	0.005	-0.004	0.001	-0.003	0.001
5	-0.218	0.006	0.001	0.002	0.001	0.002
6	-0.219	0.006	0.000	0.002	0.000	0.002
7	-0.186	0.018	0.018	0.011	0.028	0.011
8	-0.188	0.018	0.013	0.012	0.021	0.011
9	-0.197	0.020	0.018	0.013	0.030	0.013
10	-0.200	0.020	0.014	0.014	0.022	0.014
11	-0.213	0.022	0.021	0.015	0.032	0.015
12	-0.214	0.023	0.018	0.017	0.024	0.016
13	-0.138	0.041	0.051	0.033	0.081	0.032
14	-0.142	0.042	0.037	0.035	0.058	0.034
15	-0.156	0.044	0.049	0.037	0.082	0.035
16	-0.160	0.046	0.040	0.039	0.061	0.038
17	-0.205	0.052	0.060	0.045	0.082	0.042
18	-0.205	0.055	0.054	0.048	0.064	0.047
19	-0.184	0.004	-0.008	0.001	-0.007	0.001
20	-0.184	0.005	-0.008	0.001	-0.008	0.001
21	-0.187	0.005	-0.004	0.001	-0.003	0.001
22	-0.188	0.005	-0.005	0.002	-0.004	0.002
23	-0.190	0.006	0.000	0.002	0.000	0.002
24	-0.192	0.006	-0.001	0.002	-0.001	0.002
25	-0.158	0.018	0.017	0.011	0.027	0.011
26	-0.161	0.018	0.011	0.012	0.019	0.011
27	-0.170	0.020	0.016	0.013	0.028	0.013
28	-0.173	0.021	0.012	0.014	0.019	0.014
29	-0.186	0.022	0.019	0.016	0.031	0.015
30	-0.189	0.024	0.016	0.017	0.019	0.017
31	-0.110	0.041	0.049	0.033	0.078	0.032
32	-0.117	0.042	0.033	0.035	0.054	0.034
33	-0.129	0.044	0.047	0.037	0.078	0.035
34	-0.135	0.046	0.036	0.039	0.057	0.038
35	-0.179	0.052	0.057	0.045	0.078	0.042
36	-0.182	0.055	0.050	0.049	0.060	0.047
37	-0.125	0.005	-0.009	0.002	-0.008	0.002
38	-0.126	0.005	-0.009	0.002	-0.009	0.002
39	-0.128	0.005	-0.005	0.002	-0.004	0.002
40	-0.130	0.006	-0.006	0.003	-0.006	0.003
41	-0.132	0.006	-0.001	0.003	-0.002	0.003
42	-0.135	0.007	-0.003	0.004	-0.003	0.004
43	-0.100	0.018	0.014	0.012	0.023	0.012
44	-0.105	0.019	0.006	0.013	0.013	0.013
45	-0.112	0.021	0.013	0.014	0.023	0.014
46	-0.118	0.022	0.007	0.016	0.013	0.015
47	-0.129	0.023	0.016	0.017	0.026	0.016
48	-0.135	0.025	0.011	0.019	0.014	0.018
49	-0.053	0.042	0.045	0.035	0.073	0.033
50	-0.062	0.044	0.025	0.036	0.045	0.035
51	-0.072	0.045	0.042	0.038	0.072	0.036
52	-0.082	0.048	0.028	0.040	0.048	0.039
53	-0.123	0.053	0.052	0.046	0.070	0.044
54	-0.132	0.058	0.041	0.050	0.048	0.049

Table 84: Simulation 1: Mean Bias (SE) in Estimated IBE FDA Metric for Sample Size 24 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	RIS Bias	RIS SE
Complete Data Set						
1	0.000	0.000	0.000	0.000	0.001	0.000
2	0.000	0.000	0.000	0.000	0.001	0.000
3	0.000	0.000	0.000	0.000	0.002	0.000
4	0.001	0.000	0.001	0.000	0.001	0.000
5	0.000	0.000	0.000	0.000	0.004	0.000
6	0.001	0.000	0.001	0.000	0.001	0.000
7	-0.009	0.001	-0.009	0.001	-0.002	0.001
8	-0.008	0.001	-0.008	0.001	-0.006	0.001
9	0.003	0.003	0.002	0.003	0.027	0.003
10	0.006	0.004	0.006	0.004	0.006	0.004
11	0.004	0.006	0.003	0.006	0.049	0.005
12	0.010	0.006	0.009	0.006	0.007	0.006
13	0.007	0.006	0.006	0.006	0.046	0.006
14	0.014	0.007	0.013	0.007	0.017	0.007
15	0.008	0.010	0.006	0.010	0.074	0.008
16	0.018	0.010	0.016	0.010	0.018	0.010
17	0.011	0.019	0.008	0.019	0.151	0.015
18	0.030	0.019	0.027	0.019	0.020	0.019
19	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.002	0.000
22	0.001	0.000	0.001	0.000	0.001	0.000
23	0.000	0.001	0.000	0.001	0.003	0.000
24	0.001	0.001	0.001	0.001	0.001	0.001
25	-0.010	0.002	-0.010	0.002	-0.002	0.001
26	-0.008	0.002	-0.008	0.002	-0.007	0.002
27	0.002	0.004	0.001	0.004	0.026	0.003
28	0.006	0.004	0.005	0.004	0.006	0.004
29	0.003	0.006	0.002	0.006	0.048	0.005
30	0.010	0.006	0.009	0.007	0.007	0.006
31	0.006	0.006	0.005	0.007	0.044	0.006
32	0.013	0.007	0.012	0.007	0.016	0.007
33	0.007	0.010	0.005	0.010	0.073	0.008
34	0.018	0.010	0.016	0.010	0.018	0.010
35	0.010	0.019	0.006	0.019	0.149	0.015
36	0.030	0.019	0.027	0.019	0.019	0.019
37	-0.001	0.001	-0.001	0.001	0.000	0.001
38	0.000	0.001	0.000	0.001	0.000	0.001
39	-0.001	0.001	-0.001	0.001	0.001	0.001
40	0.000	0.001	0.000	0.001	0.000	0.001
41	-0.001	0.001	-0.001	0.001	0.003	0.001
42	0.001	0.002	0.001	0.002	0.001	0.002
43	-0.012	0.003	-0.012	0.003	-0.004	0.003
44	-0.009	0.003	-0.009	0.003	-0.007	0.003
45	0.000	0.004	-0.001	0.004	0.024	0.004
46	0.006	0.005	0.005	0.005	0.006	0.005
47	0.001	0.007	0.000	0.007	0.046	0.006
48	0.009	0.008	0.008	0.008	0.006	0.008
49	0.003	0.008	0.001	0.008	0.041	0.007
50	0.012	0.008	0.011	0.008	0.015	0.008
51	0.004	0.011	0.002	0.011	0.070	0.009
52	0.017	0.012	0.015	0.012	0.017	0.012
53	0.007	0.019	0.003	0.019	0.146	0.016
54	0.029	0.021	0.026	0.021	0.019	0.021

Table 84: Simulation 1: Mean Bias (SE) in Estimated IBE FDA Metric for Sample Size 24 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	RIS Bias	RIS SE
Substantial Missing Data						
1	0.186	0.001	0.001	0.000	0.001	0.000
2	0.186	0.001	0.001	0.000	0.001	0.000
3	0.187	0.001	0.001	0.000	0.003	0.000
4	0.188	0.001	0.001	0.000	0.001	0.000
5	0.188	0.001	0.001	0.000	0.005	0.000
6	0.190	0.002	0.002	0.001	0.002	0.001
7	0.193	0.003	-0.008	0.002	0.001	0.001
8	0.196	0.003	-0.006	0.002	-0.003	0.002
9	0.212	0.005	0.007	0.004	0.035	0.003
10	0.218	0.006	0.011	0.005	0.015	0.004
11	0.221	0.008	0.011	0.007	0.062	0.006
12	0.231	0.009	0.016	0.008	0.019	0.007
13	0.254	0.009	0.018	0.008	0.063	0.007
14	0.264	0.010	0.024	0.008	0.035	0.008
15	0.265	0.013	0.022	0.012	0.098	0.010
16	0.280	0.014	0.031	0.013	0.043	0.012
17	0.293	0.023	0.032	0.022	0.189	0.017
18	0.321	0.025	0.048	0.023	0.055	0.022
19	0.204	0.001	0.000	0.000	0.001	0.000
20	0.205	0.001	0.000	0.000	0.000	0.000
21	0.205	0.001	0.000	0.000	0.002	0.000
22	0.206	0.001	0.001	0.001	0.001	0.001
23	0.206	0.001	0.000	0.001	0.004	0.001
24	0.208	0.002	0.001	0.001	0.001	0.001
25	0.210	0.003	-0.009	0.002	0.000	0.002
26	0.213	0.004	-0.007	0.002	-0.005	0.002
27	0.229	0.005	0.006	0.004	0.033	0.004
28	0.235	0.006	0.009	0.005	0.013	0.005
29	0.238	0.008	0.009	0.007	0.060	0.006
30	0.247	0.009	0.014	0.008	0.017	0.008
31	0.271	0.009	0.016	0.008	0.060	0.007
32	0.281	0.010	0.022	0.009	0.032	0.008
33	0.281	0.013	0.020	0.012	0.095	0.010
34	0.296	0.014	0.028	0.013	0.040	0.012
35	0.309	0.023	0.029	0.022	0.185	0.017
36	0.337	0.025	0.044	0.024	0.051	0.023
37	0.243	0.001	0.000	0.001	0.000	0.001
38	0.243	0.001	0.000	0.001	0.000	0.001
39	0.243	0.001	-0.001	0.001	0.001	0.001
40	0.244	0.002	0.000	0.001	0.000	0.001
41	0.244	0.002	-0.001	0.002	0.003	0.001
42	0.245	0.003	-0.001	0.002	-0.001	0.002
43	0.247	0.004	-0.011	0.003	-0.003	0.003
44	0.250	0.005	-0.010	0.004	-0.008	0.004
45	0.265	0.006	0.004	0.005	0.029	0.005
46	0.271	0.007	0.006	0.006	0.009	0.006
47	0.273	0.009	0.006	0.008	0.055	0.007
48	0.283	0.011	0.009	0.010	0.012	0.009
49	0.306	0.011	0.012	0.009	0.055	0.008
50	0.316	0.012	0.018	0.011	0.027	0.010
51	0.315	0.014	0.015	0.013	0.089	0.011
52	0.330	0.016	0.022	0.015	0.033	0.014
53	0.342	0.024	0.024	0.023	0.177	0.019
54	0.369	0.027	0.036	0.026	0.042	0.024

Table 85: Simulation 1: Mean Bias (SE) in Estimated PBE FDA Metric for Sample Size 24 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE
Complete Data Set						
1	-0.006	0.001	-0.006	0.001	-0.006	0.001
2	-0.006	0.001	-0.006	0.001	-0.006	0.001
3	-0.002	0.001	-0.002	0.001	-0.002	0.001
4	-0.003	0.001	-0.003	0.001	-0.003	0.001
5	0.000	0.001	0.000	0.001	0.000	0.001
6	0.000	0.001	0.000	0.001	-0.001	0.001
7	0.003	0.007	0.003	0.007	0.003	0.007
8	-0.001	0.008	-0.001	0.008	-0.001	0.008
9	0.004	0.008	0.004	0.008	0.003	0.008
10	-0.001	0.009	-0.001	0.009	-0.002	0.009
11	0.005	0.009	0.005	0.009	0.004	0.009
12	-0.001	0.011	-0.001	0.011	-0.001	0.011
13	0.009	0.021	0.009	0.021	0.009	0.021
14	-0.003	0.023	-0.003	0.023	-0.003	0.023
15	0.011	0.023	0.010	0.023	0.010	0.023
16	-0.004	0.025	-0.004	0.025	-0.004	0.025
17	0.016	0.027	0.015	0.027	0.012	0.027
18	-0.001	0.031	-0.002	0.031	-0.005	0.031
19	-0.007	0.001	-0.007	0.001	-0.007	0.001
20	-0.007	0.001	-0.007	0.001	-0.007	0.001
21	-0.003	0.001	-0.003	0.001	-0.003	0.001
22	-0.003	0.001	-0.003	0.001	-0.003	0.001
23	0.000	0.001	0.000	0.001	0.000	0.001
24	0.000	0.001	-0.001	0.001	-0.001	0.001
25	0.002	0.007	0.002	0.007	0.002	0.007
26	-0.001	0.008	-0.001	0.008	-0.001	0.008
27	0.003	0.008	0.003	0.008	0.003	0.008
28	-0.002	0.009	-0.002	0.009	-0.002	0.009
29	0.004	0.010	0.004	0.010	0.003	0.010
30	-0.001	0.011	-0.001	0.011	-0.002	0.011
31	0.008	0.021	0.007	0.021	0.007	0.021
32	-0.004	0.023	-0.004	0.023	-0.004	0.023
33	0.009	0.023	0.009	0.023	0.008	0.023
34	-0.004	0.025	-0.005	0.025	-0.005	0.025
35	0.014	0.027	0.013	0.027	0.011	0.027
36	-0.001	0.031	-0.002	0.031	-0.006	0.031
37	-0.007	0.001	-0.007	0.001	-0.007	0.001
38	-0.007	0.001	-0.007	0.001	-0.007	0.001
39	-0.003	0.001	-0.003	0.001	-0.003	0.001
40	-0.003	0.002	-0.003	0.002	-0.003	0.002
41	-0.001	0.002	-0.001	0.002	-0.001	0.002
42	-0.001	0.002	-0.001	0.002	-0.001	0.002
43	0.000	0.008	0.000	0.008	0.000	0.008
44	-0.002	0.008	-0.002	0.008	-0.002	0.008
45	0.001	0.009	0.001	0.009	0.001	0.009
46	-0.002	0.010	-0.002	0.010	-0.002	0.010
47	0.002	0.010	0.002	0.010	0.001	0.010
48	-0.001	0.012	-0.002	0.012	-0.002	0.012
49	0.005	0.021	0.004	0.021	0.004	0.021
50	-0.005	0.023	-0.005	0.023	-0.005	0.023
51	0.006	0.023	0.006	0.023	0.005	0.023
52	-0.005	0.026	-0.006	0.026	-0.006	0.026
53	0.011	0.028	0.010	0.028	0.008	0.028
54	-0.002	0.032	-0.003	0.032	-0.006	0.032

Table 85: Simulation 1: Mean Bias (SE) in Estimated PBE FDA Metric for Sample Size 24 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE
Substantial Missing Data						
1	-0.122	0.003	-0.006	0.001	-0.006	0.001
2	-0.122	0.003	-0.006	0.001	-0.006	0.001
3	-0.123	0.003	-0.002	0.001	-0.002	0.001
4	-0.123	0.003	-0.002	0.001	-0.002	0.001
5	-0.126	0.004	0.001	0.001	0.001	0.001
6	-0.125	0.004	0.001	0.001	0.001	0.001
7	-0.116	0.012	0.009	0.008	0.010	0.008
8	-0.118	0.012	0.007	0.008	0.007	0.008
9	-0.124	0.013	0.011	0.009	0.011	0.009
10	-0.127	0.014	0.009	0.010	0.008	0.010
11	-0.136	0.015	0.013	0.010	0.012	0.010
12	-0.138	0.016	0.012	0.012	0.011	0.012
13	-0.105	0.029	0.028	0.023	0.030	0.023
14	-0.113	0.029	0.021	0.024	0.021	0.024
15	-0.120	0.031	0.030	0.025	0.030	0.025
16	-0.129	0.032	0.024	0.027	0.023	0.027
17	-0.159	0.036	0.038	0.030	0.034	0.030
18	-0.167	0.039	0.035	0.033	0.033	0.033
19	-0.104	0.003	-0.006	0.001	-0.006	0.001
20	-0.104	0.003	-0.006	0.001	-0.006	0.001
21	-0.105	0.003	-0.002	0.001	-0.002	0.001
22	-0.105	0.003	-0.003	0.001	-0.003	0.001
23	-0.108	0.004	0.000	0.001	0.000	0.001
24	-0.107	0.004	0.000	0.002	0.000	0.002
25	-0.098	0.012	0.008	0.008	0.009	0.008
26	-0.100	0.012	0.006	0.008	0.006	0.008
27	-0.107	0.013	0.010	0.009	0.009	0.009
28	-0.109	0.014	0.007	0.010	0.007	0.010
29	-0.119	0.015	0.012	0.011	0.010	0.011
30	-0.121	0.016	0.010	0.012	0.009	0.012
31	-0.089	0.029	0.026	0.023	0.027	0.023
32	-0.096	0.030	0.019	0.024	0.019	0.024
33	-0.104	0.031	0.028	0.025	0.028	0.025
34	-0.112	0.032	0.021	0.027	0.020	0.027
35	-0.144	0.036	0.036	0.030	0.030	0.030
36	-0.152	0.039	0.031	0.034	0.029	0.034
37	-0.065	0.003	-0.007	0.001	-0.007	0.001
38	-0.065	0.003	-0.007	0.001	-0.007	0.001
39	-0.067	0.003	-0.003	0.001	-0.003	0.001
40	-0.067	0.004	-0.004	0.002	-0.004	0.002
41	-0.070	0.004	-0.001	0.002	-0.001	0.002
42	-0.070	0.005	-0.002	0.002	-0.002	0.002
43	-0.061	0.012	0.006	0.008	0.007	0.008
44	-0.063	0.013	0.004	0.009	0.004	0.009
45	-0.071	0.014	0.007	0.010	0.006	0.010
46	-0.073	0.015	0.004	0.011	0.003	0.011
47	-0.084	0.016	0.008	0.011	0.006	0.011
48	-0.086	0.017	0.005	0.013	0.004	0.013
49	-0.054	0.029	0.022	0.024	0.023	0.023
50	-0.061	0.030	0.014	0.025	0.014	0.025
51	-0.070	0.032	0.024	0.026	0.022	0.026
52	-0.078	0.033	0.016	0.028	0.014	0.028
53	-0.111	0.037	0.030	0.031	0.024	0.031
54	-0.119	0.040	0.023	0.035	0.021	0.035

Table 86: Simulation 1: Mean Bias (SE) in Estimated IBE FDA Metric for Sample Size 34 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	RIS Bias	RIS SE
Complete Data Set						
1	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.001	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.003	0.000
6	0.001	0.000	0.001	0.000	0.001	0.000
7	-0.008	0.001	-0.008	0.001	-0.002	0.001
8	-0.007	0.001	-0.007	0.001	-0.006	0.001
9	0.002	0.003	0.002	0.003	0.022	0.002
10	0.004	0.003	0.004	0.003	0.004	0.003
11	0.003	0.005	0.003	0.005	0.041	0.004
12	0.007	0.005	0.007	0.005	0.006	0.005
13	0.004	0.005	0.004	0.005	0.037	0.005
14	0.009	0.006	0.008	0.006	0.011	0.006
15	0.005	0.008	0.005	0.008	0.061	0.007
16	0.012	0.008	0.012	0.008	0.012	0.008
17	0.008	0.015	0.008	0.015	0.127	0.012
18	0.021	0.016	0.021	0.016	0.017	0.016
19	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.001	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	0.003	0.000
24	0.001	0.001	0.001	0.001	0.001	0.001
25	-0.009	0.001	-0.009	0.001	-0.002	0.001
26	-0.007	0.001	-0.007	0.001	-0.006	0.001
27	0.002	0.003	0.001	0.003	0.022	0.003
28	0.004	0.003	0.004	0.003	0.004	0.003
29	0.002	0.005	0.002	0.005	0.041	0.004
30	0.007	0.005	0.007	0.005	0.006	0.005
31	0.004	0.005	0.003	0.005	0.036	0.005
32	0.009	0.006	0.008	0.006	0.010	0.006
33	0.005	0.008	0.004	0.008	0.061	0.007
34	0.012	0.009	0.011	0.009	0.012	0.009
35	0.007	0.016	0.007	0.015	0.126	0.012
36	0.021	0.016	0.021	0.016	0.017	0.016
37	0.000	0.001	0.000	0.001	0.000	0.001
38	0.000	0.001	0.000	0.001	0.000	0.001
39	0.000	0.001	0.000	0.001	0.001	0.001
40	0.000	0.001	0.000	0.001	0.000	0.001
41	0.000	0.001	0.000	0.001	0.003	0.001
42	0.001	0.001	0.001	0.001	0.001	0.001
43	-0.009	0.002	-0.009	0.002	-0.003	0.002
44	-0.008	0.003	-0.008	0.003	-0.007	0.003
45	0.001	0.004	0.001	0.004	0.021	0.003
46	0.004	0.004	0.004	0.004	0.004	0.004
47	0.002	0.006	0.001	0.006	0.040	0.005
48	0.007	0.006	0.007	0.006	0.006	0.006
49	0.003	0.006	0.002	0.006	0.035	0.006
50	0.008	0.007	0.008	0.007	0.010	0.007
51	0.003	0.009	0.003	0.009	0.059	0.008
52	0.012	0.010	0.011	0.010	0.012	0.010
53	0.006	0.016	0.006	0.016	0.125	0.013
54	0.021	0.017	0.021	0.017	0.017	0.017

Table 86: Simulation 1: Mean Bias (SE) in Estimated IBE FDA Metric for Sample Size 34 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	RIS Bias	RIS SE
Substantial Missing Data						
1	0.115	0.000	0.000	0.000	0.001	0.000
2	0.115	0.000	0.000	0.000	0.000	0.000
3	0.116	0.001	0.000	0.000	0.002	0.000
4	0.116	0.001	0.001	0.000	0.001	0.000
5	0.117	0.001	0.000	0.000	0.004	0.000
6	0.118	0.001	0.001	0.000	0.001	0.000
7	0.117	0.002	-0.009	0.001	-0.001	0.001
8	0.119	0.002	-0.008	0.001	-0.006	0.001
9	0.132	0.004	0.003	0.003	0.024	0.003
10	0.135	0.004	0.005	0.004	0.006	0.003
11	0.137	0.006	0.004	0.006	0.044	0.005
12	0.142	0.007	0.007	0.006	0.007	0.006
13	0.157	0.007	0.007	0.006	0.042	0.005
14	0.163	0.007	0.011	0.006	0.015	0.006
15	0.164	0.010	0.008	0.009	0.067	0.008
16	0.172	0.010	0.013	0.010	0.017	0.009
17	0.182	0.018	0.011	0.018	0.135	0.014
18	0.196	0.019	0.021	0.018	0.019	0.018
19	0.128	0.001	0.000	0.000	0.000	0.000
20	0.128	0.001	0.000	0.000	0.000	0.000
21	0.129	0.001	0.000	0.000	0.001	0.000
22	0.129	0.001	0.000	0.000	0.000	0.000
23	0.129	0.001	0.000	0.001	0.003	0.000
24	0.131	0.001	0.001	0.001	0.001	0.001
25	0.129	0.002	-0.009	0.002	-0.002	0.001
26	0.131	0.002	-0.008	0.002	-0.007	0.002
27	0.144	0.004	0.002	0.003	0.023	0.003
28	0.147	0.004	0.004	0.004	0.005	0.004
29	0.149	0.006	0.003	0.006	0.043	0.005
30	0.155	0.007	0.007	0.006	0.006	0.006
31	0.169	0.007	0.005	0.006	0.041	0.005
32	0.175	0.007	0.009	0.007	0.014	0.007
33	0.176	0.010	0.006	0.009	0.066	0.008
34	0.184	0.011	0.012	0.010	0.015	0.010
35	0.193	0.018	0.009	0.018	0.133	0.014
36	0.208	0.019	0.020	0.019	0.018	0.018
37	0.155	0.001	-0.001	0.001	0.000	0.001
38	0.155	0.001	0.000	0.001	0.000	0.001
39	0.156	0.001	-0.001	0.001	0.001	0.001
40	0.156	0.001	0.000	0.001	0.000	0.001
41	0.156	0.001	0.000	0.001	0.003	0.001
42	0.158	0.002	0.000	0.002	0.000	0.002
43	0.155	0.003	-0.011	0.003	-0.004	0.003
44	0.157	0.004	-0.010	0.003	-0.008	0.003
45	0.170	0.005	0.000	0.004	0.022	0.004
46	0.174	0.005	0.003	0.005	0.004	0.005
47	0.175	0.007	0.001	0.007	0.042	0.006
48	0.181	0.008	0.005	0.007	0.005	0.007
49	0.193	0.008	0.002	0.007	0.039	0.007
50	0.200	0.009	0.007	0.008	0.011	0.008
51	0.200	0.011	0.003	0.010	0.064	0.009
52	0.210	0.012	0.010	0.011	0.013	0.011
53	0.218	0.019	0.006	0.019	0.131	0.015
54	0.234	0.021	0.018	0.020	0.016	0.019

Table 87: Simulation 1: Mean Bias (SE) in Estimated PBE FDA Metric for Sample Size 34 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE
Complete Data Set						
1	-0.005	0.000	-0.005	0.000	-0.005	0.000
2	-0.005	0.000	-0.005	0.000	-0.005	0.000
3	-0.001	0.001	-0.001	0.001	-0.001	0.001
4	-0.001	0.001	-0.001	0.001	-0.001	0.001
5	0.000	0.001	0.000	0.001	0.000	0.001
6	0.000	0.001	0.000	0.001	0.000	0.001
7	0.001	0.006	0.001	0.006	0.002	0.006
8	0.002	0.006	0.002	0.006	0.002	0.006
9	0.001	0.007	0.001	0.007	0.001	0.007
10	0.002	0.007	0.002	0.007	0.002	0.007
11	0.002	0.008	0.002	0.008	0.001	0.008
12	0.003	0.009	0.003	0.009	0.003	0.009
13	0.004	0.017	0.004	0.017	0.004	0.017
14	0.005	0.018	0.005	0.018	0.005	0.018
15	0.004	0.018	0.003	0.019	0.003	0.019
16	0.006	0.020	0.005	0.020	0.005	0.020
17	0.005	0.022	0.005	0.022	0.003	0.022
18	0.009	0.025	0.009	0.025	0.009	0.025
19	-0.005	0.001	-0.005	0.001	-0.005	0.001
20	-0.005	0.001	-0.005	0.001	-0.005	0.001
21	-0.002	0.001	-0.002	0.001	-0.002	0.001
22	-0.001	0.001	-0.001	0.001	-0.001	0.001
23	0.000	0.001	0.000	0.001	0.000	0.001
24	0.000	0.001	0.000	0.001	0.000	0.001
25	0.001	0.006	0.001	0.006	0.001	0.006
26	0.002	0.006	0.002	0.006	0.002	0.006
27	0.001	0.007	0.001	0.007	0.001	0.007
28	0.002	0.007	0.002	0.007	0.002	0.007
29	0.001	0.008	0.001	0.008	0.001	0.008
30	0.003	0.009	0.003	0.009	0.003	0.009
31	0.003	0.017	0.003	0.017	0.003	0.017
32	0.005	0.018	0.005	0.018	0.005	0.018
33	0.003	0.019	0.003	0.019	0.002	0.019
34	0.005	0.020	0.005	0.020	0.005	0.020
35	0.004	0.023	0.004	0.023	0.003	0.023
36	0.009	0.025	0.009	0.025	0.009	0.025
37	-0.005	0.001	-0.005	0.001	-0.005	0.001
38	-0.005	0.001	-0.005	0.001	-0.005	0.001
39	-0.002	0.001	-0.002	0.001	-0.002	0.001
40	-0.002	0.001	-0.002	0.001	-0.002	0.001
41	0.000	0.001	0.000	0.001	0.000	0.001
42	0.000	0.002	0.000	0.002	0.000	0.002
43	0.000	0.006	0.000	0.006	0.000	0.006
44	0.001	0.007	0.001	0.007	0.001	0.007
45	0.000	0.007	0.000	0.007	0.000	0.007
46	0.002	0.008	0.002	0.008	0.002	0.008
47	0.000	0.008	0.000	0.008	0.000	0.008
48	0.003	0.010	0.003	0.009	0.003	0.009
49	0.002	0.017	0.002	0.017	0.002	0.017
50	0.004	0.019	0.004	0.019	0.004	0.019
51	0.002	0.019	0.002	0.019	0.001	0.019
52	0.005	0.021	0.005	0.021	0.005	0.021
53	0.003	0.023	0.003	0.023	0.002	0.023
54	0.009	0.026	0.009	0.026	0.009	0.026

Table 87: Simulation 1: Mean Bias (SE) in Estimated PBE FDA Metric for Sample Size 34 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE
Substantial Missing Data						
1	-0.079	0.002	-0.005	0.000	-0.005	0.000
2	-0.079	0.002	-0.005	0.001	-0.005	0.001
3	-0.081	0.002	-0.002	0.001	-0.002	0.001
4	-0.080	0.002	-0.002	0.001	-0.002	0.001
5	-0.083	0.003	0.000	0.001	0.000	0.001
6	-0.082	0.003	0.000	0.001	0.000	0.001
7	-0.079	0.009	0.001	0.006	0.002	0.006
8	-0.078	0.009	0.001	0.006	0.001	0.006
9	-0.088	0.010	-0.001	0.007	0.000	0.007
10	-0.085	0.010	0.001	0.008	0.001	0.008
11	-0.099	0.011	-0.001	0.008	-0.001	0.008
12	-0.095	0.012	0.002	0.009	0.002	0.009
13	-0.083	0.022	0.000	0.018	0.003	0.018
14	-0.080	0.022	0.002	0.019	0.003	0.019
15	-0.098	0.023	-0.002	0.020	0.000	0.020
16	-0.093	0.024	0.002	0.022	0.003	0.022
17	-0.134	0.028	-0.004	0.024	-0.004	0.024
18	-0.123	0.030	0.007	0.027	0.007	0.027
19	-0.066	0.002	-0.005	0.001	-0.005	0.001
20	-0.066	0.002	-0.005	0.001	-0.005	0.001
21	-0.068	0.002	-0.002	0.001	-0.002	0.001
22	-0.067	0.002	-0.002	0.001	-0.002	0.001
23	-0.070	0.003	-0.001	0.001	-0.001	0.001
24	-0.069	0.003	0.000	0.001	0.000	0.001
25	-0.066	0.009	0.000	0.006	0.001	0.006
26	-0.065	0.009	0.000	0.007	0.001	0.007
27	-0.075	0.010	-0.001	0.007	-0.001	0.007
28	-0.073	0.010	0.000	0.008	0.000	0.008
29	-0.087	0.011	-0.002	0.008	-0.002	0.008
30	-0.082	0.012	0.001	0.009	0.001	0.009
31	-0.072	0.022	-0.001	0.018	0.001	0.018
32	-0.068	0.022	0.001	0.019	0.002	0.019
33	-0.086	0.023	-0.003	0.020	-0.001	0.020
34	-0.081	0.024	0.001	0.022	0.002	0.022
35	-0.122	0.028	-0.005	0.024	-0.006	0.024
36	-0.111	0.030	0.006	0.027	0.006	0.027
37	-0.039	0.002	-0.006	0.001	-0.006	0.001
38	-0.039	0.002	-0.006	0.001	-0.006	0.001
39	-0.041	0.002	-0.003	0.001	-0.003	0.001
40	-0.040	0.002	-0.002	0.001	-0.002	0.001
41	-0.043	0.003	-0.001	0.002	-0.001	0.002
42	-0.042	0.003	-0.001	0.002	0.000	0.002
43	-0.041	0.009	-0.002	0.007	-0.001	0.007
44	-0.039	0.009	-0.001	0.007	-0.001	0.007
45	-0.050	0.010	-0.003	0.008	-0.003	0.008
46	-0.046	0.010	-0.001	0.009	-0.001	0.009
47	-0.061	0.012	-0.004	0.009	-0.004	0.009
48	-0.056	0.012	0.000	0.010	0.000	0.010
49	-0.047	0.022	-0.004	0.019	-0.002	0.019
50	-0.043	0.023	-0.001	0.020	-0.001	0.020
51	-0.062	0.024	-0.006	0.020	-0.004	0.020
52	-0.055	0.025	-0.001	0.022	0.000	0.022
53	-0.097	0.029	-0.008	0.025	-0.009	0.025
54	-0.085	0.031	0.003	0.028	0.003	0.028

Table 88: Simulation 1: Mean Bias (SE) in Estimated IBE FDA Metric for Sample Size 80 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	RIS Bias	RIS SE
Complete Data Set						
1	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.001	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.002	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000
7	-0.006	0.001	-0.006	0.001	-0.002	0.001
8	-0.006	0.001	-0.006	0.001	-0.006	0.001
9	0.000	0.002	0.000	0.002	0.013	0.002
10	0.001	0.002	0.001	0.002	0.001	0.002
11	-0.001	0.003	-0.001	0.003	0.025	0.003
12	0.001	0.004	0.001	0.004	0.001	0.003
13	0.000	0.004	0.000	0.004	0.021	0.003
14	0.002	0.004	0.002	0.004	0.002	0.004
15	-0.001	0.005	-0.001	0.005	0.035	0.005
16	0.002	0.006	0.002	0.006	0.002	0.006
17	-0.002	0.010	-0.002	0.010	0.077	0.008
18	0.004	0.011	0.004	0.011	0.003	0.011
19	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.001	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	0.002	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000
25	-0.006	0.001	-0.006	0.001	-0.002	0.001
26	-0.006	0.001	-0.006	0.001	-0.006	0.001
27	0.000	0.002	0.000	0.002	0.013	0.002
28	0.001	0.002	0.001	0.002	0.000	0.002
29	-0.001	0.003	-0.001	0.003	0.025	0.003
30	0.001	0.004	0.001	0.004	0.001	0.004
31	0.000	0.004	0.000	0.004	0.020	0.003
32	0.001	0.004	0.001	0.004	0.001	0.004
33	-0.001	0.005	-0.001	0.005	0.035	0.005
34	0.002	0.006	0.002	0.006	0.001	0.006
35	-0.003	0.010	-0.002	0.010	0.077	0.008
36	0.003	0.011	0.004	0.011	0.002	0.011
37	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.001	0.000	0.001	0.001	0.000
40	0.000	0.001	0.000	0.001	0.000	0.001
41	0.000	0.001	0.000	0.001	0.002	0.001
42	0.000	0.001	0.000	0.001	0.000	0.001
43	-0.007	0.001	-0.007	0.001	-0.003	0.001
44	-0.007	0.002	-0.007	0.002	-0.006	0.002
45	-0.001	0.002	-0.001	0.002	0.012	0.002
46	0.000	0.003	0.000	0.003	0.000	0.003
47	-0.001	0.004	-0.001	0.004	0.024	0.003
48	0.001	0.004	0.001	0.004	0.001	0.004
49	-0.001	0.004	-0.001	0.004	0.020	0.004
50	0.000	0.005	0.000	0.005	0.000	0.005
51	-0.002	0.006	-0.002	0.006	0.035	0.005
52	0.001	0.006	0.001	0.006	0.000	0.006
53	-0.003	0.011	-0.003	0.011	0.076	0.009
54	0.002	0.011	0.003	0.011	0.001	0.011

Table 88: Simulation 1: Mean Bias (SE) in Estimated IBE FDA Metric for Sample Size 80 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	RIS Bias	RIS SE
Substantial Missing Data						
1	0.042	0.000	0.000	0.000	0.000	0.000
2	0.042	0.000	0.000	0.000	0.000	0.000
3	0.042	0.000	0.000	0.000	0.001	0.000
4	0.042	0.000	0.000	0.000	0.000	0.000
5	0.042	0.000	0.000	0.000	0.002	0.000
6	0.042	0.000	0.000	0.000	0.000	0.000
7	0.039	0.001	-0.006	0.001	-0.002	0.001
8	0.039	0.001	-0.006	0.001	-0.006	0.001
9	0.046	0.002	0.000	0.002	0.014	0.002
10	0.047	0.002	0.002	0.002	0.002	0.002
11	0.047	0.004	0.000	0.004	0.026	0.003
12	0.049	0.004	0.003	0.004	0.002	0.004
13	0.054	0.004	0.001	0.004	0.022	0.003
14	0.056	0.004	0.003	0.004	0.003	0.004
15	0.055	0.006	0.001	0.006	0.038	0.005
16	0.058	0.006	0.004	0.006	0.004	0.006
17	0.059	0.011	0.000	0.011	0.081	0.009
18	0.066	0.011	0.008	0.011	0.007	0.011
19	0.047	0.000	0.000	0.000	0.000	0.000
20	0.047	0.000	0.000	0.000	0.000	0.000
21	0.047	0.000	0.000	0.000	0.001	0.000
22	0.047	0.000	0.000	0.000	0.000	0.000
23	0.048	0.000	0.000	0.000	0.002	0.000
24	0.048	0.000	0.000	0.000	0.000	0.000
25	0.044	0.001	-0.006	0.001	-0.002	0.001
26	0.044	0.001	-0.006	0.001	-0.006	0.001
27	0.051	0.002	0.000	0.002	0.014	0.002
28	0.052	0.002	0.001	0.002	0.001	0.002
29	0.052	0.004	0.000	0.004	0.026	0.003
30	0.055	0.004	0.002	0.004	0.002	0.004
31	0.059	0.004	0.001	0.004	0.022	0.003
32	0.061	0.004	0.003	0.004	0.003	0.004
33	0.060	0.006	0.000	0.006	0.038	0.005
34	0.063	0.006	0.004	0.006	0.004	0.006
35	0.064	0.011	0.000	0.011	0.080	0.009
36	0.071	0.012	0.008	0.011	0.006	0.011
37	0.059	0.000	0.000	0.000	0.000	0.000
38	0.059	0.001	0.000	0.000	0.000	0.000
39	0.059	0.001	0.000	0.001	0.001	0.001
40	0.059	0.001	0.000	0.001	0.000	0.001
41	0.059	0.001	0.000	0.001	0.002	0.001
42	0.059	0.001	0.000	0.001	0.000	0.001
43	0.055	0.002	-0.006	0.002	-0.002	0.001
44	0.055	0.002	-0.006	0.002	-0.006	0.002
45	0.063	0.003	0.000	0.003	0.014	0.002
46	0.063	0.003	0.001	0.003	0.001	0.003
47	0.064	0.004	0.000	0.004	0.026	0.003
48	0.065	0.005	0.002	0.004	0.001	0.004
49	0.070	0.005	0.001	0.004	0.022	0.004
50	0.071	0.005	0.002	0.005	0.002	0.005
51	0.071	0.006	0.000	0.006	0.038	0.005
52	0.074	0.007	0.003	0.007	0.003	0.007
53	0.075	0.011	0.000	0.011	0.080	0.009
54	0.080	0.012	0.007	0.012	0.005	0.012

Table 89: Simulation 1: Mean Bias (SE) in Estimated PBE FDA Metric for Sample Size 80 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE
Complete Data Set						
1	-0.003	0.000	-0.003	0.000	-0.003	0.000
2	-0.003	0.000	-0.003	0.000	-0.003	0.000
3	-0.001	0.000	-0.001	0.000	-0.001	0.000
4	-0.001	0.001	-0.001	0.001	-0.001	0.001
5	0.000	0.001	0.000	0.001	0.000	0.001
6	0.000	0.001	0.000	0.001	0.000	0.001
7	-0.003	0.004	-0.003	0.004	-0.003	0.004
8	-0.004	0.004	-0.004	0.004	-0.004	0.004
9	-0.004	0.004	-0.004	0.004	-0.004	0.004
10	-0.004	0.005	-0.004	0.005	-0.004	0.005
11	-0.004	0.005	-0.004	0.005	-0.004	0.005
12	-0.004	0.006	-0.004	0.006	-0.004	0.006
13	-0.010	0.011	-0.010	0.011	-0.010	0.011
14	-0.010	0.012	-0.010	0.012	-0.010	0.012
15	-0.010	0.012	-0.010	0.012	-0.010	0.012
16	-0.011	0.013	-0.011	0.013	-0.011	0.013
17	-0.011	0.015	-0.010	0.015	-0.011	0.015
18	-0.011	0.016	-0.011	0.016	-0.011	0.016
19	-0.003	0.000	-0.003	0.000	-0.003	0.000
20	-0.003	0.000	-0.003	0.000	-0.003	0.000
21	-0.001	0.000	-0.001	0.000	-0.001	0.000
22	-0.001	0.001	-0.001	0.001	-0.001	0.001
23	-0.001	0.001	0.000	0.001	-0.001	0.001
24	-0.001	0.001	-0.001	0.001	-0.001	0.001
25	-0.004	0.004	-0.004	0.004	-0.004	0.004
26	-0.004	0.004	-0.004	0.004	-0.004	0.004
27	-0.004	0.004	-0.004	0.004	-0.004	0.004
28	-0.004	0.005	-0.004	0.005	-0.004	0.005
29	-0.004	0.005	-0.004	0.005	-0.004	0.005
30	-0.004	0.006	-0.004	0.006	-0.004	0.006
31	-0.010	0.011	-0.010	0.011	-0.010	0.011
32	-0.011	0.012	-0.011	0.012	-0.011	0.012
33	-0.011	0.012	-0.010	0.012	-0.011	0.012
34	-0.011	0.013	-0.011	0.013	-0.011	0.013
35	-0.011	0.015	-0.011	0.015	-0.011	0.015
36	-0.012	0.016	-0.012	0.016	-0.012	0.016
37	-0.003	0.000	-0.003	0.000	-0.003	0.000
38	-0.003	0.001	-0.003	0.001	-0.003	0.001
39	-0.001	0.001	-0.001	0.001	-0.001	0.001
40	-0.001	0.001	-0.001	0.001	-0.001	0.001
41	-0.001	0.001	-0.001	0.001	-0.001	0.001
42	-0.001	0.001	-0.001	0.001	-0.001	0.001
43	-0.004	0.004	-0.004	0.004	-0.004	0.004
44	-0.004	0.004	-0.004	0.004	-0.004	0.004
45	-0.004	0.005	-0.004	0.005	-0.004	0.005
46	-0.005	0.005	-0.005	0.005	-0.005	0.005
47	-0.004	0.005	-0.004	0.005	-0.005	0.005
48	0.001	0.006	0.001	0.006	0.001	0.006
49	-0.011	0.011	-0.011	0.011	-0.011	0.011
50	-0.012	0.012	-0.012	0.012	-0.012	0.012
51	-0.011	0.012	-0.011	0.012	-0.011	0.012
52	-0.012	0.014	-0.012	0.014	-0.012	0.014
53	-0.012	0.015	-0.012	0.015	-0.012	0.015
54	-0.013	0.017	-0.013	0.017	-0.013	0.017

Table 89: Simulation 1: Mean Bias (SE) in Estimated PBE FDA Metric for Sample Size 80 (1000 runs per simulation)

Sim	MoM Bias	MoM SE	UN Bias	UN SE	CSH Bias	CSH SE
Substantial Missing Data						
1	-0.031	0.001	-0.003	0.000	-0.003	0.000
2	-0.031	0.001	-0.003	0.000	-0.003	0.000
3	-0.032	0.001	-0.001	0.000	-0.001	0.000
4	-0.032	0.001	-0.001	0.001	-0.001	0.001
5	-0.033	0.001	0.000	0.001	0.000	0.001
6	-0.033	0.001	-0.001	0.001	-0.001	0.001
7	-0.036	0.004	-0.003	0.004	-0.003	0.004
8	-0.037	0.005	-0.004	0.004	-0.004	0.004
9	-0.040	0.005	-0.004	0.004	-0.004	0.004
10	-0.040	0.005	-0.004	0.005	-0.004	0.005
11	-0.044	0.006	-0.004	0.005	-0.004	0.005
12	-0.045	0.006	-0.005	0.006	-0.005	0.006
13	-0.046	0.012	-0.010	0.011	-0.009	0.011
14	-0.048	0.013	-0.010	0.012	-0.010	0.012
15	-0.052	0.013	-0.010	0.012	-0.010	0.012
16	-0.054	0.014	-0.011	0.013	-0.011	0.013
17	-0.068	0.016	-0.011	0.015	-0.012	0.015
18	-0.068	0.017	-0.013	0.017	-0.013	0.017
19	-0.026	0.001	-0.003	0.000	-0.003	0.000
20	-0.026	0.001	-0.003	0.000	-0.003	0.000
21	-0.026	0.001	-0.001	0.001	-0.001	0.001
22	-0.027	0.001	-0.001	0.001	-0.001	0.001
23	-0.028	0.001	0.000	0.001	-0.001	0.001
24	-0.028	0.001	-0.001	0.001	-0.001	0.001
25	-0.031	0.005	-0.003	0.004	-0.003	0.004
26	-0.031	0.005	-0.004	0.004	-0.004	0.004
27	-0.034	0.005	-0.004	0.005	-0.004	0.004
28	-0.035	0.005	-0.004	0.005	-0.004	0.005
29	-0.039	0.006	-0.004	0.005	-0.004	0.005
30	-0.040	0.006	-0.005	0.006	-0.005	0.006
31	-0.041	0.012	-0.010	0.011	-0.009	0.011
32	-0.043	0.013	-0.011	0.012	-0.011	0.012
33	-0.047	0.013	-0.010	0.012	-0.010	0.012
34	-0.049	0.014	-0.012	0.013	-0.012	0.013
35	-0.063	0.016	-0.012	0.015	-0.012	0.015
36	-0.064	0.017	-0.014	0.017	-0.014	0.017
37	-0.014	0.001	-0.003	0.001	-0.003	0.001
38	-0.014	0.001	-0.003	0.001	-0.003	0.001
39	-0.015	0.001	-0.001	0.001	-0.001	0.001
40	-0.015	0.001	-0.001	0.001	-0.001	0.001
41	-0.016	0.001	-0.001	0.001	-0.001	0.001
42	-0.017	0.001	-0.001	0.001	-0.001	0.001
43	-0.019	0.005	-0.003	0.004	-0.003	0.004
44	-0.020	0.005	-0.004	0.004	-0.004	0.004
45	-0.023	0.005	-0.004	0.005	-0.004	0.005
46	-0.025	0.006	-0.005	0.005	-0.005	0.005
47	-0.028	0.006	-0.004	0.006	-0.004	0.006
48	-0.030	0.007	-0.006	0.006	-0.006	0.006
49	-0.030	0.012	-0.010	0.012	-0.009	0.012
50	-0.032	0.013	-0.012	0.012	-0.012	0.012
51	-0.036	0.013	-0.010	0.013	-0.010	0.013
52	-0.039	0.014	-0.013	0.014	-0.013	0.014
53	-0.052	0.016	-0.012	0.015	-0.012	0.015
54	-0.054	0.018	-0.015	0.017	-0.015	0.017

Table 90: Simulation 1: Percentage of IBE Failures for Sample Size 16 (1000 runs per simulation)

Sim	MoM	UN	RIS
Complete Data Set			
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	5.7	0.4	0.3
7	80.3	78	68.7
8	94	92.9	88.7
9	42.4	57.5	57
10	75.9	86	88.4
11	18.2	30.4	47.5
12	63.4	72.6	89
13	68.7	82.3	97.7
14	88.1	94.8	99.5
15	42.5	58.4	91.7
16	75.9	86	98.1
17	15.8	26.1	73.8
18	62.3	70.6	94.6
19	3	1.6	2.1
20	10	5.1	5.2
21	8.7	5.3	5.6
22	38.8	25.5	25.6
23	15.7	11.1	9.9
24	63.5	49	46.4
25	95.7	97.3	95.3
26	98.9	99.1	98.4
27	62.9	79.8	87.3
28	86.2	93.9	96.5
29	30.6	49.7	73.5
30	71.2	82.1	93.9
31	75.5	89	99
32	92.5	96.4	99.7
33	50.2	67.4	94.8
34	80.4	89.4	98.7
35	19.6	31.1	81.1
36	63.1	74.6	96
37	100	100	100
38	100	100	100
39	100	100	100
40	100	100	100
41	100	100	100
42	100	100	100
43	100	100	100
44	100	100	100
45	100	100	100
46	99.8	100	100
47	95.4	98.9	100
48	98.3	99.4	100
49	98.3	99.8	100
50	99.6	99.9	100
51	91.3	97	100
52	96.9	99	99.9

Mom: Method-of-Moment Estimation, CF Inference
UN: Unstructured REML Estimation, Asymptotic Inference
RIS: Constrained REML Estimation, Asymptotic Inference

Table 90: Simulation 1: Percentage of IBE Failures for Sample Size 16 (1000 runs per simulation)

Sim	MoM	UN	RIS
53	57.7	77.2	99.2
54	84.8	92.3	99.4
Substantial Missing Data			
1	100	0.1	0
2	100	0.2	0
3	100	0.1	0
4	100	2.3	1.6
5	100	1.3	0.5
6	100	12.5	9
7	100	87.3	83.2
8	100	93.5	91.8
9	100	86.3	78.8
10	100	93.3	90.6
11	98.4	71.2	72.5
12	98.6	88.2	88.4
13	99.3	93.5	97.7
14	99.8	96.5	98.4
15	97.7	87.1	95.3
16	98.4	93.3	97.9
17	84.4	68.4	87.2
18	92.2	87.2	94.6
19	100	18.5	22.3
20	100	27	30
21	100	26.6	31.4
22	100	45.7	48.2
23	100	36.2	36.5
24	100	63.4	60.1
25	100	95.1	93.9
26	100	98.3	96.7
27	100	90.7	88.3
28	100	96.1	93.8
29	99.1	79.7	83.2
30	99.1	91.3	92.2
31	99.5	93.9	98.4
32	99.9	97.5	99.3
33	98	88.6	96.5
34	98.7	94.8	98.9
35	85.6	71.1	89
36	93.9	88.3	95.1
37	100	99.5	100
38	100	99.6	100
39	100	100	100
40	100	100	100
41	100	99.8	100
42	100	100	100
43	100	99.8	100
44	100	99.9	100
45	100	100	100
46	100	99.8	100
47	100	98.5	99.9
48	100	99	99.8
49	100	99.2	100
50	100	99.1	100

Mom: Method-of-Moment Estimation, CF Inference
UN: Unstructured REML Estimation, Asymptotic Inference
RIS: Constrained REML Estimation, Asymptotic Inference

Table 90: Simulation 1: Percentage of IBE Failures for Sample Size 16 (1000 runs per simulation)

Sim	MoM	UN	RIS
51	99.3	97.4	100
52	99.9	98.8	99.9
53	94.1	88.7	98.7
54	97.2	94.4	98.3
Mom: Method-of-Moment Estimation, CF Inference UN: Unstructured REML Estimation, Asymptotic Inference RIS: Constrained REML Estimation, Asymptotic Inference			

Table 91: Simulation 1: Percentage of IBE Failures for Sample Size 24 (1000 runs per simulation)

Sim	MoM	UN	RIS
Complete Data Set			
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	68.5	67.3	56.2
8	91.3	91.5	84.4
9	19.5	29.9	37.8
10	63	70.7	87.6
11	5	7.2	27.7
12	43.2	51.1	88.3
13	52	62.3	97.4
14	81.4	87.6	99.9
15	19.5	29.9	88.7
16	63	70.7	98.8
17	4.1	6.1	59.6
18	41	48.2	96.1
19	0.2	0.1	0.2
20	1.8	1.1	1.3
21	1.5	0.6	0.8
22	20	13.7	13.9
23	3.8	2.9	2.3
24	47.4	35.3	32.9
25	95.2	97.5	96.1
26	99.2	99.6	98.8
27	47.7	61.9	81.9
28	80.2	87	96.3
29	12.8	20.1	62.7
30	57.9	66.2	94.2
31	64	76	99.1
32	87.6	91.6	100
33	29.3	41.8	94.4
34	70.4	77.7	93.3
35	5.5	8.6	70.9
36	45.3	53.4	96.6
37	100	100	100
38	100	100	100
39	100	100	100
40	100	100	100
41	100	100	100
42	100	100	100
43	100	100	100
44	100	100	100
45	99.9	100	100
46	99.9	100	100
47	95.2	98.6	100
48	98.4	99.6	100
49	99.4	99.8	100
50	99.9	99.9	100
51	89.2	95.9	100
52	96	98.3	100

Mom: Method-of-Moment Estimation, CF Inference
UN: Unstructured REML Estimation, Asymptotic Inference
RIS: Constrained REML Estimation, Asymptotic Inference

Table 91: Simulation 1: Percentage of IBE Failures for Sample Size 24 (1000 runs per simulation)

Sim	MoM	UN	RIS
53	40.3	55.9	99.1
54	77.4	84.5	99.9
Substantial Missing Data			
1	100	0	0
2	100	0	0
3	100	0	0
4	100	0	0
5	100	0	0
6	100	1.1	0.6
7	100	77.5	67.8
8	100	91.8	86
9	95.6	51	54.7
10	98.1	80.5	87.9
11	75.5	24.2	45.3
12	90.1	65.1	88.6
13	93.1	76.1	97.3
14	97.7	91.6	99.6
15	78.2	51.4	91.6
16	90.4	80.5	98.5
17	40.9	21.4	71.3
18	74.7	63.2	94.3
19	100	2.3	2.6
20	100	6.5	6.9
21	100	5	4.9
22	100	25.3	25
23	100	10.2	7.9
24	100	46.1	43.3
25	100	96.7	95
26	100	98.5	97.6
27	98.5	73.2	86.2
28	99.6	90.8	96.2
29	85.1	42.2	70
30	94	76.5	93.8
31	95.6	84	98.3
32	98.3	95	99.7
33	83.3	60.1	95.4
34	93.3	85	98.8
35	47.2	25.7	77.9
36	77.7	66.5	95.1
37	100	100	100
38	100	100	100
39	100	100	100
40	100	100	100
41	100	100	100
42	100	100	100
43	100	100	100
44	100	100	100
45	100	99.9	100
46	100	99.8	100
47	99.8	99.4	100
48	99.8	99.6	100
49	99.9	99.8	100
50	99.9	99.8	99.9

Mom: Method-of-Moment Estimation, CF Inference
UN: Unstructured REML Estimation, Asymptotic Inference
RIS: Constrained REML Estimation, Asymptotic Inference

Table 91: Simulation 1: Percentage of IBE Failures for Sample Size 24 (1000 runs per simulation)

Sim	MoM	UN	RIS
51	98.2	96.9	100
52	99.3	98.9	99.9
53	76.1	69.8	99.6
54	91.1	89.1	99.5
Mom: Method-of-Moment Estimation, CF Inference UN: Unstructured REML Estimation, Asymptotic Inference RIS: Constrained REML Estimation, Asymptotic Inference			

Table 92: Simulation 1: Percentage of IBE Failures for Sample Size 34 (1000 runs per simulation)

Sim	MoM	UN	RIS
Complete Data Set			
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	55.2	52.8	42.3
8	89.1	88.2	81.3
9	8.7	12.1	22.7
10	48.8	55.9	85
11	1.3	1.5	14.1
12	27.1	31.5	85.5
13	34.7	42.1	97.1
14	73.6	80	99.7
15	8.7	12.1	85.5
16	48.8	55.9	98.5
17	1	1.1	46.2
18	24.4	28.8	96.5
19	0.1	0	0
20	0.3	0.1	0.1
21	0.1	0.1	0.1
22	9.2	5.5	5.3
23	0.3	0.2	0.2
24	31.4	25.1	24.1
25	94.7	96.7	95.7
26	98.6	99.8	99.4
27	31.1	41.9	77.6
28	73.4	79.3	96.9
29	5.1	6.7	50.1
30	43.1	49.2	94.2
31	51.7	61.8	99.3
32	82.8	88.4	100
33	15.5	21.1	93.7
34	58.7	66.3	99.2
35	1.6	1.9	60.1
36	30	35.6	97.7
37	100	100	100
38	100	100	100
39	100	100	100
40	100	100	100
41	100	100	100
42	100	100	100
43	100	100	100
44	100	100	100
45	100	100	100
46	100	100	100
47	95	98.2	100
48	99.2	99.9	100
49	99.1	99.8	100
50	100	100	100
51	86.7	92.9	100
52	96.7	98.1	100

Mom: Method-of-Moment Estimation, CF Inference
UN: Unstructured REML Estimation, Asymptotic Inference
RIS: Constrained REML Estimation, Asymptotic Inference

Table 92: Simulation 1: Percentage of IBE Failures for Sample Size 34 (1000 runs per simulation)

Sim	MoM	UN	RIS
53	25	36.3	99.7
54	69.9	77.8	100
Substantial Missing Data			
1	100	0	0
2	100	0	0
3	100	0	0
4	100	0	0
5	99.9	0	0
6	99.8	0.1	0
7	99.6	63.7	53.7
8	99.9	90.2	85.5
9	81.5	22.8	35.5
10	94.5	64.2	85.6
11	40.4	4	23.6
12	74	43.7	86.2
13	82.6	55.8	97.1
14	94.8	84.3	99.4
15	50.7	22.8	88
16	79.1	64.2	98.2
17	11.1	3.5	55.6
18	52.3	41.5	95.5
19	100	0.1	0.1
20	100	0.9	0.9
21	100	0.6	0.6
22	100	10.3	10.1
23	100	1.3	1.4
24	100	31.3	29.7
25	99.9	95.3	95.3
26	100	99.5	98.8
27	92.6	54.5	80.4
28	98.1	81.9	95.8
29	57.8	15.4	58.4
30	83.6	59.2	92.6
31	88.3	69.1	98.8
32	97.3	91	100
33	61	33.6	94.1
34	85.6	72.1	98.7
35	15.6	5.8	67.7
36	56.7	46.5	96.3
37	100	100	100
38	100	100	100
39	100	100	100
40	100	100	100
41	100	100	100
42	100	100	100
43	100	100	100
44	100	100	100
45	100	100	100
46	100	99.9	100
47	99.6	97.3	100
48	99.7	99.4	100
49	99.9	99.7	100
50	99.9	100	100

Mom: Method-of-Moment Estimation, CF Inference
UN: Unstructured REML Estimation, Asymptotic Inference
RIS: Constrained REML Estimation, Asymptotic Inference

Table 92: Simulation 1: Percentage of IBE Failures for Sample Size 34 (1000 runs per simulation)

Sim	MoM	UN	RIS
51	95.8	93	100
52	99	97.7	100
53	58.3	49.2	99.4
54	84.7	80.1	100
Mom: Method-of-Moment Estimation, CF Inference UN: Unstructured REML Estimation, Asymptotic Inference RIS: Constrained REML Estimation, Asymptotic Inference			

Table 93: Simulation 1: Percentage of IBE Failures for Sample Size 80 (1000 runs per simulation)

Sim	MoM	UN	RIS
Complete Data Set			
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	16.4	14.1	8.8
8	80.3	80.4	72.5
9	0	0	1
10	15.3	17.2	76.8
11	0	0	0.1
12	2.3	2.6	77.9
13	5.3	6.4	95.4
14	48.6	52.9	100
15	0	0	64
16	15.3	17.2	99.6
17	0	0	12.4
18	1.6	1.8	96.4
19	0	0	0
20	0	0	0
21	0	0	0
22	0	0	0
23	0	0	0
24	3.2	2.4	2.2
25	94.5	95.7	96.1
26	99.2	99.5	99.8
27	4.6	6.4	52.4
28	48.2	53.3	98.3
29	0	0	15.1
30	10.7	12	96
31	17	20.7	99.8
32	67.2	71.5	100
33	0.3	0.8	87.8
34	25.4	33.1	99.8
35	0	0	25.6
36	2.9	3.3	98.5
37	100	100	100
38	100	100	100
39	100	100	100
40	100	100	100
41	100	100	100
42	100	100	100
43	100	100	100
44	100	100	100
45	100	100	100
46	100	100	100
47	95.2	97.2	100
48	99.6	99.6	100
49	99.7	99.9	100
50	100	100	100
51	78.4	84.4	100
52	97.7	98.4	100

Mom: Method-of-Moment Estimation, CF Inference
UN: Unstructured REML Estimation, Asymptotic Inference
RIS: Constrained REML Estimation, Asymptotic Inference

Table 93: Simulation 1: Percentage of IBE Failures for Sample Size 80 (1000 runs per simulation)

Sim	MoM	UN	RIS
53	2.4	4.3	99.9
54	43.5	48.1	100
Substantial Missing Data			
1	0	0	0
2	0	0	0
3	0	0	0
4	0.6	0	0
5	0	0	0
6	6.1	0	0
7	91	18.3	14.1
8	99	82.4	74.3
9	8	0.1	2.4
10	50.1	21.4	76.6
11	0.2	0	0.8
12	14.4	4.1	78.8
13	23.3	9.8	96.4
14	70.2	55.2	99.9
15	1.2	0.1	67.6
16	30.5	21.4	99.5
17	0	0	16
18	5.2	3.3	96.7
19	100	0	0
20	100	0	0
21	99.2	0	0
22	99.8	0.1	0
23	87.9	0	0
24	98.4	2.8	3.2
25	99.7	95.6	95.9
26	99.9	99.4	99.8
27	36.7	9.1	57.3
28	81.3	55.6	98.4
29	1.1	0	19.6
30	30.4	15.2	95.4
31	44.2	25.6	99.7
32	85.4	73.7	99.9
33	4.9	1.4	89.3
34	43.5	30.5	99.8
35	0	0	30.6
36	9	5.2	98.5
37	100	100	100
38	100	100	100
39	100	100	100
40	100	100	100
41	100	100	100
42	100	100	100
43	100	100	100
44	100	100	100
45	100	100	100
46	100	100	100
47	98.2	97.6	100
48	99.8	99.7	100
49	99.8	99.8	100
50	100	100	100

Mom: Method-of-Moment Estimation, CF Inference
UN: Unstructured REML Estimation, Asymptotic Inference
RIS: Constrained REML Estimation, Asymptotic Inference

Table 93: Simulation 1: Percentage of IBE Failures for Sample Size 80 (1000 runs per simulation)

Sim	MoM	UN	RIS
51	89.6	84.8	100
52	98.5	98	100
53	8.4	6.1	99.9
54	55.6	51.9	100
Mom: Method-of-Moment Estimation, CF Inference UN: Unstructured REML Estimation, Asymptotic Inference RIS: Constrained REML Estimation, Asymptotic Inference			

Table 94: Simulation 1: Percentage of PBE Failures for Sample Size 16 (1000 runs per simulation)

Sim	MoM	UN	CSH
Complete Data Set			
1	31.6	0	0
2	32	0.4	0.4
3	14.8	0	0
4	21.3	1.5	1.5
5	5.3	0.2	0.3
6	11.7	3.1	3.1
7	47.9	4.2	4.6
8	48.1	8.2	8.6
9	24.5	1.8	2.4
10	31.9	8.3	8.4
11	12	0.8	1.3
12	19.1	7.6	7.7
13	37.9	2.7	3.7
14	40.6	8.5	8.7
15	24.5	1.8	2.4
16	31.9	8.3	8.4
17	10.1	0.7	1.1
18	17.1	6.9	7
19	97.9	83	85.1
20	97.8	85.4	86.5
21	91.8	76.8	79.8
22	88.1	83.2	83.9
23	65.3	57.8	61.2
24	67	68.1	68.4
25	61.8	11.2	14
26	58.4	20.2	20.9
27	37.5	6.6	8.1
28	39.8	17.4	17.9
29	17.3	3.4	4.1
30	26.2	14.6	14.8
31	44.1	4.9	5.9
32	44	12	12.2
33	28.8	3.3	4.5
34	35.5	11.6	11.8
35	11.8	1.3	2.1
36	19.7	8.9	9.1
37	100	100	100
38	100	100	100
39	100	100	100
40	100	100	100
41	100	100	99.8
42	100	100	99.9
43	99.5	97.6	97.7
44	98.8	97	97.1
45	95.7	91.2	91.9
46	92.3	91	91.1
47	81.4	75.8	78.9
48	78.5	79.7	79.9
49	77.2	43.9	47.3
50	73.7	54.4	55
51	63.4	32.9	35.7
52	62.5	46.8	47.3

Mom: Method-of-Moment Estimation, CF Inference
UN: Unstructured REML Estimation, Asymptotic Inference
CSH: Constrained REML Estimation, Asymptotic Inference

Table 94: Simulation 1: Percentage of PBE Failures for Sample Size 16 (1000 runs per simulation)

Sim	MoM	UN	CSH
53	32.8	16.2	18.2
54	39.9	32.5	32.8
Substantial Missing Data			
1	100	5.9	6.9
2	100	9.4	10
3	100	7	8.6
4	99.9	14.8	16.5
5	99.8	5.7	7.2
6	99.5	19	19.2
7	89.2	23.8	30
8	87.8	33.1	36.9
9	82.9	20	25.2
10	81.4	31.5	35.2
11	73	14.3	18.1
12	72.5	27.3	29.3
13	75.6	22.9	28.5
14	73.8	32.5	36.6
15	69	19.9	25
16	68.7	31.5	35.1
17	53.8	13.8	16.2
18	53.7	26.5	27.6
19	100	83.7	89
20	100	84.8	89.6
21	100	79.3	85
22	100	83.2	87.2
23	100	69.9	74.8
24	99.9	75.1	76.1
25	91.6	35.7	44.4
26	92.3	43.7	49.5
27	87.1	30.5	38.1
28	84.5	42.6	45.8
29	77.3	23.1	27.6
30	75.5	36.2	38.2
31	78	27	33.3
32	77.1	36.6	40.4
33	70.9	22.6	28.2
34	71.4	36	39.2
35	56	15.2	18.4
36	54.3	29.3	30
37	100	99.5	99.6
38	100	99.6	99.5
39	100	100	99.4
40	100	100	99.8
41	100	99.8	98.2
42	100	100	98.1
43	99.4	95.8	97.7
44	99.3	95	96.8
45	98.3	89.8	93.6
46	97.2	91.2	92.7
47	94.8	80.4	85.9
48	92.8	84.8	86.8
49	89.9	61.8	70.4
50	87.8	67.7	71.9

Mom: Method-of-Moment Estimation, CF Inference
UN: Unstructured REML Estimation, Asymptotic Inference
CSH: Constrained REML Estimation, Asymptotic Inference

Table 94: Simulation 1: Percentage of PBE Failures for Sample Size 16 (1000 runs per simulation)

Sim	MoM	UN	CSH
51	84.2	54.6	62.7
52	82.2	64.4	68.8
53	69.9	40.6	47
54	68.5	54.7	56.5
Mom: Method-of-Moment Estimation, CF Inference UN: Unstructured REML Estimation, Asymptotic Inference CSH: Constrained REML Estimation, Asymptotic Inference			

Table 95: Simulation 1: Percentage of PBE Failures for Sample Size 24 (1000 runs per simulation)

Sim	MoM	UN	CSH
Complete Data Set			
1	4.4	0	0
2	5.7	0	0
3	0.9	0	0
4	3.4	0.1	0.1
5	0.2	0	0
6	1.5	0.3	0.3
7	10.1	0.1	0.1
8	14	1.1	1.1
9	3.1	0	0
10	7.7	1.6	1.6
11	0.7	0	0
12	3.8	0.8	0.8
13	6.5	0.1	0.1
14	11.6	1.2	1.4
15	3.1	0	0
16	7.7	1.6	1.6
17	0.4	0	0
18	3.2	0.8	0.8
19	96.7	75.3	77.1
20	95.4	79.1	79.4
21	82.4	64	66.7
22	81	72.5	72.8
23	43.2	33	35.6
24	49.1	48.8	48.9
25	20.6	1	1.5
26	24.8	4.5	4.7
27	7.4	0.3	0.4
28	14.2	3.8	3.8
29	2.4	0.2	0.2
30	7	2.7	2.8
31	9.2	0.2	0.3
32	14	1.9	2.2
33	4.6	0	0.1
34	9.8	2.4	2.5
35	0.6	0	0
36	4	1.2	1.3
37	100	100	100
38	100	100	100
39	100	100	100
40	100	100	100
41	100	100	100
42	100	100	99.8
43	98.5	92.6	93.1
44	96.6	91.6	91.6
45	89.5	78.8	80
46	85.6	82.1	82.3
47	65.5	55.9	57.9
48	67.3	66.1	66.1
49	48.1	17.2	18.1
50	48.3	26.1	26.1
51	28.7	10.3	11.3
52	36.2	21.1	21.2

Mom: Method-of-Moment Estimation, CF Inference
UN: Unstructured REML Estimation, Asymptotic Inference
CSH: Constrained REML Estimation, Asymptotic Inference

Table 95: Simulation 1: Percentage of PBE Failures for Sample Size 24 (1000 runs per simulation)

Sim	MoM	UN	CSH
53	8.4	3.9	4.1
54	17.2	12.2	12.2
Substantial Missing Data			
1	98.9	0	0
2	98.2	0.1	0.1
3	96.6	0.1	0.2
4	94.7	1	1
5	89.2	0.1	0.1
6	85.7	1.4	1.4
7	61	2.2	2.6
8	60.1	5.5	5.7
9	45.3	1.1	1.3
10	46.6	4.8	4.8
11	28.6	0.6	0.7
12	32.9	3.8	3.8
13	38.8	1.9	1.9
14	39.8	5.4	5.5
15	28.1	1.1	1.3
16	29.8	4.8	4.8
17	10.7	0.5	0.5
18	15.8	3.9	3.9
19	100	79.7	82.5
20	100	82.6	83.9
21	99.7	69.7	73.5
22	99.4	76.5	77.1
23	97.9	44.7	47.9
24	95.8	58.6	58.8
25	70.2	5.8	6.2
26	69.2	12	12.3
27	56.4	4.2	4.5
28	55.2	10.5	10.6
29	36.1	2.3	2.6
30	40	7.5	7.6
31	42.6	2.9	3
32	43.7	7.1	7.4
33	31.9	2	2.2
34	34.2	6.7	6.7
35	13.9	0.5	0.6
36	18.6	4.6	4.6
37	100	100	100
38	100	100	100
39	100	100	99.9
40	100	100	100
41	100	100	99.9
42	100	100	99.7
43	98.1	91.9	92.2
44	97.6	92.8	93
45	93.4	82.3	83.8
46	91.4	84.8	85
47	83.2	65.4	68.4
48	80.5	71.7	72.4
49	70.4	27.6	31.8
50	70.6	38.1	38.7

Mom: Method-of-Moment Estimation, CF Inference
UN: Unstructured REML Estimation, Asymptotic Inference
CSH: Constrained REML Estimation, Asymptotic Inference

Table 95: Simulation 1: Percentage of PBE Failures for Sample Size 24 (1000 runs per simulation)

Sim	MoM	UN	CSH
51	60	19.7	22.4
52	60.6	32.6	33.2
53	32.8	9.5	10.3
54	37.1	22.3	22.5
Mom: Method-of-Moment Estimation, CF Inference UN: Unstructured REML Estimation, Asymptotic Inference CSH: Constrained REML Estimation, Asymptotic Inference			

Table 96: Simulation 1: Percentage of PBE Failures for Sample Size 34 (1000 runs per simulation)

Sim	MoM	UN	CSH
Complete Data Set			
1	0.1	0	0
2	0.3	0	0
3	0.1	0	0
4	0.2	0	0
5	0	0	0
6	0.1	0	0
7	0.5	0	0
8	1.9	0	0
9	0.1	0	0
10	1.3	0	0
11	0	0	0
12	0.6	0.1	0.1
13	0.2	0	0
14	1.7	0	0
15	0.1	0	0
16	1.3	0	0
17	0	0	0
18	0.6	0.1	0.1
19	94.1	67.9	69.7
20	93.7	71.3	72
21	71.5	48.5	50.6
22	71.2	59.1	59.1
23	24.3	16.9	18.6
24	32.4	31.7	31.8
25	3.1	0	0
26	6.8	0.1	0.1
27	0.4	0	0
28	3	0.3	0.4
29	0.1	0	0
30	1.5	0.5	0.5
31	0.6	0	0
32	2.2	0	0
33	0.1	0	0
34	1.7	0	0
35	0	0	0
36	0.8	0.1	0.1
37	100	100	100
38	100	100	100
39	100	100	100
40	100	100	100
41	100	100	100
42	100	100	100
43	96.2	86.4	86.9
44	95	88.1	88.2
45	80.6	66.8	68.2
46	80.6	72.4	72.4
47	47.9	37.8	39.6
48	52	48.9	49
49	19.9	4.5	5.2
50	24.7	11.4	11.4
51	9.5	1.6	1.9
52	16.4	7.8	7.8

Mom: Method-of-Moment Estimation, CF Inference
UN: Unstructured REML Estimation, Asymptotic Inference
CSH: Constrained REML Estimation, Asymptotic Inference

Table 96: Simulation 1: Percentage of PBE Failures for Sample Size 34 (1000 runs per simulation)

Sim	MoM	UN	CSH
53	1.2	0.2	0.2
54	5.7	4	4
Substantial Missing Data			
1	85.8	0	0
2	86.7	0	0
3	73.6	0	0
4	72.2	0	0
5	51.3	0	0
6	52.5	0.1	0.1
7	24.6	0	0
8	25.8	0.2	0.2
9	12.1	0	0
10	14.6	0.3	0.3
11	4.9	0	0
12	6.1	0.4	0.4
13	9.2	0	0
14	10.6	0.1	0.1
15	5.3	0	0
16	6.7	0.3	0.3
17	0.7	0	0
18	2.4	0.4	0.4
19	99.3	72	74.2
20	99.4	74.2	74.8
21	97.2	53.7	55.8
22	96.3	64.7	64.9
23	85.9	23.6	25.3
24	85.8	38.7	38.8
25	37.2	0.2	0.3
26	37.6	1.5	1.5
27	20.1	0.1	0.1
28	22.2	1.9	1.9
29	7.7	0	0
30	11.1	1.1	1.1
31	11.7	0.1	0.1
32	14.1	0.3	0.3
33	6.3	0	0
34	9.1	0.6	0.6
35	1.2	0	0
36	2.4	0.7	0.7
37	100	100	100
38	100	100	100
39	100	100	100
40	100	100	100
41	100	100	100
42	100	100	100
43	96.6	87.3	88
44	96.1	88.1	88.2
45	86.9	69.8	71.9
46	86.5	76.2	76.2
47	66.2	43.2	45.5
48	65.6	55.3	55.4
49	43.1	8.6	9.4
50	45.2	15.9	15.9

Mom: Method-of-Moment Estimation, CF Inference
UN: Unstructured REML Estimation, Asymptotic Inference
CSH: Constrained REML Estimation, Asymptotic Inference

Table 96: Simulation 1: Percentage of PBE Failures for Sample Size 34 (1000 runs per simulation)

Sim	MoM	UN	CSH
51	28.7	5.1	5.5
52	32.1	11.8	11.9
53	8.9	1.1	1.2
54	13.7	6.9	6.9
Mom: Method-of-Moment Estimation, CF Inference UN: Unstructured REML Estimation, Asymptotic Inference CSH: Constrained REML Estimation, Asymptotic Inference			

Table 97: Simulation 1: Percentage of PBE Failures for Sample Size 80 (1000 runs per simulation)

Sim	MoM	UN	CSH
Complete Data Set			
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	0	0	0
9	0	0	0
10	0	0	0
11	0	0	0
12	0	0	0
13	0	0	0
14	0	0	0
15	0	0	0
16	0	0	0
17	0	0	0
18	0	0	0
19	81	41.3	42
20	79.4	47.2	47.4
21	31.3	12.2	12.8
22	36	23.7	23.7
23	0.7	0.5	0.5
24	4.7	4.2	4.2
25	0	0	0
26	0	0	0
27	0	0	0
28	0	0	0
29	0	0	0
30	0	0	0
31	0	0	0
32	0	0	0
33	0	0	0
34	0	0	0
35	0	0	0
36	0	0	0
37	100	100	100
38	100	100	100
39	100	100	100
40	100	100	100
41	100	100	100
42	100	100	100
43	86.4	65.4	66.5
44	83.2	68.5	68.5
45	45.7	27.2	28.3
46	48.5	37.7	37.7
47	8.3	4	4.4
48	16	12.4	12.4
49	0.1	0	0
50	0.5	0	0
51	0	0	0
52	0.1	0	0

Mom: Method-of-Moment Estimation, CF Inference
UN: Unstructured REML Estimation, Asymptotic Inference
CSH: Constrained REML Estimation, Asymptotic Inference

Table 97: Simulation 1: Percentage of PBE Failures for Sample Size 80 (1000 runs per simulation)

Sim	MoM	UN	CSH
53	0	0	0
54	0	0	0
Substantial Missing Data			
1	5.3	0	0
2	5	0	0
3	1	0	0
4	1.4	0	0
5	0.2	0	0
6	0.2	0	0
7	0	0	0
8	0	0	0
9	0	0	0
10	0	0	0
11	0	0	0
12	0	0	0
13	0	0	0
14	0	0	0
15	0	0	0
16	0	0	0
17	0	0	0
18	0	0	0
19	88.8	44.3	45.3
20	88.9	50.7	50.9
21	61.1	14.2	15.7
22	60	26.3	26.3
23	19.2	1	1
24	23.2	4.4	4.4
25	0.1	0	0
26	0.1	0	0
27	0	0	0
28	0	0	0
29	0	0	0
30	0	0	0
31	0	0	0
32	0	0	0
33	0	0	0
34	0	0	0
35	0	0	0
36	0	0	0
37	100	100	100
38	100	100	100
39	100	100	100
40	100	100	100
41	100	100	100
42	100	100	100
43	87.1	67.1	68
44	85.3	70.5	70.6
45	55.7	29.6	30.6
46	54.7	38.4	38.4
47	14.7	4.8	5.1
48	20.4	13.7	13.7
49	0.7	0	0
50	1.5	0	0

Mom: Method-of-Moment Estimation, CF Inference
UN: Unstructured REML Estimation, Asymptotic Inference
CSH: Constrained REML Estimation, Asymptotic Inference

Table 97: Simulation 1: Percentage of PBE Failures for Sample Size 80 (1000 runs per simulation)

Sim	MoM	UN	CSH
51	0.1	0	0
52	0.3	0	0
53	0	0	0
54	0.1	0	0

Mom: Method-of-Moment Estimation, CF Inference
UN: Unstructured REML Estimation, Asymptotic Inference
CSH: Constrained REML Estimation, Asymptotic Inference

Table 98: Findings from Bridging Simulations (500 Runs per Simulation)

Sim	δ Bias	δ SE	$\frac{\sigma_T^2}{\sigma_R^2}$ Bias	$\frac{\sigma_T^2}{\sigma_R^2}$ SE	FDA Bias	FDA SE	KLD Bias	KLD SE
1	-0.005	0.001	0.058	0.017	0.168	0.019	0.59	0.051
2	-0.005	0.001	0.058	0.017	1.646	0.323	6.863	0.789
3	-0.005	0.001	0.058	0.017	4.969	0.871	19.8	2.154
4	-0.007	0.002	0.117	0.035	0.297	0.038	0.448	0.028
5	-0.007	0.002	0.117	0.035	1.647	0.35	3.804	0.442
6	-0.007	0.002	0.117	0.035	4.878	0.899	10.99	1.167
7	-0.011	0.003	0.351	0.105	0.817	0.113	0.57	0.054
8	-0.011	0.003	0.351	0.105	1.846	0.447	1.909	0.282
9	-0.011	0.003	0.351	0.105	4.85	1.01	5.22	0.649
10	-0.041	0.011	0.059	0.017	0.168	0.019	0.59	0.051
11	-0.041	0.011	0.059	0.017	0.147	0.027	0.606	0.064
12	-0.041	0.011	0.059	0.017	0.167	0.039	0.75	0.092
13	-0.054	0.014	0.117	0.035	0.297	0.038	0.448	0.028
14	-0.054	0.014	0.117	0.035	0.26	0.044	0.432	0.038
15	-0.054	0.014	0.117	0.035	0.268	0.057	0.496	0.056
16	-0.086	0.023	0.351	0.105	0.817	0.113	0.57	0.054
17	-0.086	0.023	0.351	0.105	0.739	0.118	0.523	0.058
18	-0.086	0.023	0.351	0.105	0.719	0.13	0.527	0.068
19	-0.001	0.001	0.021	0.01	0.078	0.011	0.266	0.024
20	-0.001	0.001	0.021	0.01	0.669	0.19	2.953	0.396
21	-0.001	0.001	0.021	0.01	1.899	0.496	8.336	1.077
22	-0.002	0.001	0.042	0.02	0.137	0.022	0.202	0.01
23	-0.002	0.001	0.042	0.02	0.665	0.21	1.615	0.226
24	-0.002	0.001	0.042	0.02	1.852	0.519	4.554	0.588
25	-0.004	0.002	0.127	0.06	0.368	0.063	0.256	0.029
26	-0.004	0.002	0.127	0.06	0.743	0.279	0.789	0.165
27	-0.004	0.002	0.127	0.06	1.819	0.602	2.083	0.36
28	-0.01	0.008	0.021	0.01	0.079	0.011	0.266	0.024
29	-0.01	0.008	0.021	0.01	0.078	0.018	0.286	0.033
30	-0.01	0.008	0.021	0.01	0.09	0.026	0.355	0.048
31	-0.017	0.011	0.042	0.02	0.137	0.022	0.202	0.01
32	-0.017	0.011	0.042	0.02	0.128	0.028	0.203	0.02
33	-0.017	0.011	0.042	0.02	0.135	0.037	0.233	0.031
34	-0.033	0.017	0.127	0.06	0.368	0.063	0.256	0.029
35	-0.033	0.017	0.127	0.06	0.341	0.07	0.239	0.034
36	-0.033	0.017	0.127	0.06	0.334	0.08	0.241	0.041

Table 99: Parametric and Nonparametric Bootstrapped Lower and Upper Bounds for δ and $\frac{\sigma_T^2}{\sigma_R^2}$ (500 Runs per Simulation)

Sim	Mean P	Mean NP	SE P	SE NP	Med P	Med NP	5th Q P	5th Q NP	95th Q P	95th Q NP
Lower Bound of δ										
1	-0.06	-0.05	0.03	0.03	-0.06	-0.05	-0.11	-0.11	0.00	0.00
2	0.35	0.35	0.03	0.03	0.35	0.35	0.29	0.29	0.40	0.40
3	0.63	0.64	0.03	0.03	0.63	0.64	0.58	0.58	0.69	0.69
4	-0.08	-0.07	0.04	0.04	-0.08	-0.07	-0.15	-0.14	-0.01	0.00
5	0.33	0.34	0.04	0.04	0.33	0.34	0.26	0.27	0.40	0.41
6	0.62	0.62	0.04	0.04	0.62	0.63	0.55	0.56	0.69	0.70
7	-0.13	-0.11	0.07	0.07	-0.12	-0.11	-0.24	-0.22	-0.01	0.01
8	0.28	0.30	0.07	0.07	0.28	0.30	0.16	0.19	0.40	0.41
9	0.57	0.59	0.07	0.07	0.57	0.59	0.45	0.47	0.69	0.70
10	-0.48	-0.44	0.26	0.26	-0.48	-0.43	-0.91	-0.89	-0.04	-0.01
11	-0.08	-0.03	0.26	0.26	-0.07	-0.03	-0.51	-0.48	0.37	0.40
12	0.21	0.25	0.26	0.26	0.22	0.26	-0.22	-0.19	0.65	0.69
13	-0.62	-0.55	0.34	0.33	-0.60	-0.54	-1.17	-1.09	-0.05	0.02
14	-0.22	-0.15	0.34	0.33	-0.20	-0.14	-0.76	-0.69	0.36	0.43
15	0.07	0.14	0.34	0.33	0.09	0.15	-0.48	-0.40	0.65	0.72
16	-1.01	-0.86	0.56	0.54	-1.00	-0.85	-1.93	-1.75	-0.05	0.05
17	-0.60	-0.46	0.56	0.54	-0.59	-0.44	-1.52	-1.35	0.35	0.45
18	-0.31	-0.17	0.56	0.54	-0.30	-0.15	-1.23	-1.06	0.64	0.74
19	-0.04	-0.04	0.02	0.02	-0.04	-0.04	-0.08	-0.08	0.00	0.00
20	0.37	0.37	0.02	0.02	0.36	0.37	0.33	0.33	0.41	0.41
21	0.65	0.66	0.02	0.02	0.65	0.65	0.61	0.62	0.69	0.70
22	-0.05	-0.05	0.03	0.03	-0.05	-0.05	-0.10	-0.10	0.00	0.00
23	0.35	0.36	0.03	0.03	0.35	0.36	0.30	0.31	0.41	0.41
24	0.64	0.65	0.03	0.03	0.64	0.64	0.59	0.60	0.69	0.70
25	-0.08	-0.08	0.05	0.05	-0.08	-0.08	-0.16	-0.16	0.00	0.01
26	0.32	0.33	0.05	0.05	0.32	0.33	0.24	0.25	0.41	0.41
27	0.61	0.62	0.05	0.05	0.61	0.62	0.53	0.54	0.69	0.70
28	-0.32	-0.30	0.19	0.19	-0.33	-0.31	-0.63	-0.61	0.01	0.02
29	0.09	0.10	0.19	0.19	0.07	0.09	-0.22	-0.20	0.42	0.43
30	0.38	0.39	0.19	0.19	0.36	0.38	0.06	0.09	0.70	0.71
31	-0.41	-0.38	0.24	0.24	-0.42	-0.39	-0.81	-0.78	0.01	0.03
32	0.00	0.02	0.24	0.24	-0.02	0.01	-0.40	-0.37	0.42	0.43
33	0.29	0.31	0.24	0.24	0.27	0.30	-0.11	-0.08	0.71	0.72
34	-0.66	-0.61	0.39	0.39	-0.66	-0.62	-1.29	-1.25	0.01	0.06
35	-0.25	-0.20	0.39	0.39	-0.26	-0.21	-0.89	-0.84	0.42	0.47
36	0.04	0.09	0.39	0.39	0.03	0.08	-0.60	-0.55	0.71	0.76
Upper Bound of δ										
1	0.05	0.04	0.03	0.03	0.05	0.04	0.00	-0.01	0.11	0.10
2	0.46	0.45	0.03	0.03	0.46	0.45	0.40	0.40	0.51	0.51
3	0.74	0.74	0.03	0.03	0.74	0.74	0.69	0.68	0.80	0.79
4	0.06	0.06	0.04	0.04	0.06	0.06	-0.01	-0.01	0.13	0.13
5	0.47	0.46	0.04	0.04	0.47	0.46	0.40	0.39	0.54	0.53
6	0.76	0.75	0.04	0.04	0.76	0.75	0.69	0.68	0.83	0.82
7	0.10	0.09	0.07	0.07	0.10	0.09	-0.01	-0.02	0.21	0.20
8	0.51	0.49	0.07	0.07	0.51	0.49	0.40	0.38	0.62	0.60
9	0.80	0.78	0.07	0.07	0.80	0.78	0.68	0.67	0.90	0.89
10	0.40	0.36	0.26	0.26	0.40	0.36	-0.02	-0.07	0.85	0.81
11	0.81	0.76	0.26	0.26	0.81	0.76	0.38	0.33	1.26	1.22
12	1.09	1.05	0.26	0.26	1.10	1.05	0.67	0.62	1.54	1.50
13	0.51	0.44	0.34	0.33	0.51	0.45	-0.05	-0.10	1.08	1.02
14	0.92	0.85	0.34	0.33	0.92	0.85	0.36	0.30	1.48	1.42
15	1.21	1.14	0.34	0.33	1.20	1.14	0.64	0.59	1.77	1.71
P: Parametric, NP: Nonparametric Percentile										

Table 99: Parametric and Nonparametric Bootstrapped Lower and Upper Bounds for δ and $\frac{\sigma_T^2}{\sigma_R^2}$ (500 Runs per Simulation)

Sim	Mean P	Mean NP	SE P	SE NP	Med P	Med NP	5th Q P	5th Q NP	95th Q P	95th Q NP
16	0.84	0.69	0.56	0.54	0.84	0.71	-0.07	-0.20	1.69	1.58
17	1.24	1.10	0.56	0.54	1.24	1.11	0.33	0.21	2.09	1.99
18	1.53	1.38	0.56	0.54	1.53	1.40	0.62	0.49	2.38	2.28
19	0.04	0.04	0.02	0.02	0.04	0.04	0.00	0.00	0.08	0.08
20	0.44	0.44	0.02	0.02	0.44	0.44	0.40	0.40	0.48	0.48
21	0.73	0.73	0.02	0.02	0.73	0.73	0.69	0.69	0.77	0.77
22	0.05	0.04	0.03	0.03	0.05	0.04	0.00	0.00	0.10	0.10
23	0.45	0.45	0.03	0.03	0.45	0.45	0.40	0.40	0.51	0.50
24	0.74	0.74	0.03	0.03	0.74	0.74	0.69	0.69	0.79	0.79
25	0.07	0.07	0.05	0.05	0.07	0.07	-0.01	-0.01	0.16	0.15
26	0.48	0.47	0.05	0.05	0.48	0.47	0.40	0.39	0.57	0.56
27	0.77	0.76	0.05	0.05	0.77	0.76	0.68	0.68	0.85	0.85
28	0.30	0.28	0.19	0.19	0.29	0.28	-0.01	-0.02	0.63	0.61
29	0.70	0.69	0.19	0.19	0.69	0.69	0.40	0.38	1.04	1.02
30	0.99	0.98	0.19	0.19	0.98	0.97	0.68	0.67	1.33	1.30
31	0.37	0.35	0.25	0.25	0.37	0.35	-0.03	-0.04	0.80	0.78
32	0.78	0.76	0.25	0.25	0.78	0.75	0.38	0.37	1.21	1.19
33	1.07	1.04	0.25	0.25	1.06	1.04	0.67	0.65	1.50	1.47
34	0.59	0.54	0.40	0.40	0.58	0.52	-0.08	-0.12	1.29	1.23
35	0.99	0.95	0.40	0.40	0.98	0.93	0.33	0.29	1.69	1.64
36	1.28	1.24	0.40	0.40	1.27	1.21	0.61	0.58	1.98	1.93
Lower Bound of $\frac{\sigma_T^2}{\sigma_R^2}$										
1	0.23	0.15	0.16	0.12	0.19	0.12	0.06	0.03	0.51	0.37
2	0.23	0.15	0.16	0.12	0.19	0.12	0.06	0.03	0.51	0.37
3	0.23	0.15	0.16	0.12	0.19	0.12	0.06	0.03	0.51	0.37
4	0.46	0.31	0.32	0.25	0.37	0.24	0.13	0.07	1.01	0.74
5	0.46	0.31	0.32	0.25	0.37	0.24	0.13	0.07	1.01	0.74
6	0.46	0.31	0.32	0.25	0.37	0.24	0.13	0.07	1.01	0.74
7	1.37	0.92	0.96	0.74	1.12	0.73	0.38	0.20	3.04	2.21
8	1.37	0.92	0.96	0.74	1.12	0.73	0.38	0.20	3.04	2.21
9	1.37	0.92	0.96	0.74	1.12	0.73	0.38	0.20	3.04	2.21
10	0.23	0.15	0.16	0.12	0.19	0.12	0.06	0.03	0.51	0.37
11	0.23	0.15	0.16	0.12	0.19	0.12	0.06	0.03	0.51	0.37
12	0.23	0.15	0.16	0.12	0.19	0.12	0.06	0.03	0.51	0.37
13	0.46	0.31	0.32	0.25	0.37	0.24	0.13	0.07	1.01	0.74
14	0.46	0.31	0.32	0.25	0.37	0.24	0.13	0.07	1.01	0.74
15	0.46	0.31	0.32	0.25	0.37	0.24	0.13	0.07	1.01	0.74
16	1.37	0.92	0.96	0.74	1.12	0.73	0.38	0.20	3.04	2.21
17	1.37	0.92	0.96	0.74	1.12	0.73	0.38	0.20	3.04	2.21
18	1.37	0.92	0.96	0.74	1.12	0.73	0.38	0.20	3.04	2.21
19	0.28	0.23	0.12	0.11	0.26	0.22	0.12	0.10	0.50	0.43
20	0.28	0.23	0.12	0.11	0.26	0.22	0.12	0.10	0.50	0.43
21	0.28	0.23	0.12	0.11	0.26	0.22	0.12	0.10	0.50	0.43
22	0.55	0.47	0.24	0.21	0.52	0.44	0.24	0.19	1.01	0.85
23	0.55	0.47	0.24	0.21	0.52	0.44	0.24	0.19	1.01	0.85
24	0.55	0.47	0.24	0.21	0.52	0.44	0.24	0.19	1.01	0.85
25	1.66	1.41	0.71	0.64	1.55	1.31	0.73	0.58	3.03	2.55
26	1.66	1.41	0.71	0.64	1.55	1.31	0.73	0.58	3.03	2.55
27	1.66	1.41	0.71	0.64	1.55	1.31	0.73	0.58	3.03	2.55
28	0.28	0.23	0.12	0.11	0.26	0.22	0.12	0.10	0.50	0.43
29	0.28	0.23	0.12	0.11	0.26	0.22	0.12	0.10	0.50	0.43
30	0.28	0.23	0.12	0.11	0.26	0.22	0.12	0.10	0.50	0.43
31	0.55	0.47	0.24	0.21	0.52	0.44	0.24	0.19	1.01	0.85
P: Parametric, NP: Nonparametric Percentile										

Table 99: Parametric and Nonparametric Bootstrapped Lower and Upper Bounds for δ and $\frac{\sigma_T^2}{\sigma_R^2}$ (500 Runs per Simulation)

Sim	Mean P	Mean NP	SE P	SE NP	Med P	Med NP	5th Q P	5th Q NP	95th Q P	95th Q NP
32	0.55	0.47	0.24	0.21	0.52	0.44	0.24	0.19	1.01	0.85
33	0.55	0.47	0.24	0.21	0.52	0.44	0.24	0.19	1.01	0.85
34	1.66	1.41	0.71	0.64	1.55	1.31	0.73	0.58	3.03	2.55
35	1.66	1.41	0.71	0.64	1.55	1.31	0.73	0.58	3.03	2.55
36	1.66	1.41	0.71	0.64	1.55	1.31	0.73	0.58	3.03	2.55
Upper Bound of $\frac{\sigma_T^2}{\sigma_R^2}$										
1	1.91	1.13	1.33	0.80	1.55	0.91	0.53	0.30	4.22	2.49
2	1.91	1.13	1.33	0.80	1.55	0.91	0.53	0.30	4.22	2.49
3	1.91	1.13	1.33	0.80	1.55	0.91	0.53	0.30	4.22	2.49
4	3.82	2.25	2.67	1.59	3.10	1.81	1.05	0.60	8.45	4.98
5	3.82	2.25	2.67	1.59	3.10	1.81	1.05	0.60	8.45	4.98
6	3.82	2.25	2.67	1.59	3.10	1.81	1.05	0.60	8.45	4.98
7	11.45	6.76	8.00	4.78	9.31	5.44	3.16	1.81	25.34	14.94
8	11.45	6.76	8.00	4.78	9.31	5.44	3.16	1.81	25.34	14.94
9	11.45	6.76	8.00	4.78	9.31	5.44	3.16	1.81	25.34	14.94
10	1.91	1.13	1.33	0.80	1.55	0.91	0.53	0.30	4.22	2.49
11	1.91	1.13	1.33	0.80	1.55	0.91	0.53	0.30	4.22	2.49
12	1.91	1.13	1.33	0.80	1.55	0.91	0.53	0.30	4.22	2.49
13	3.82	2.25	2.67	1.59	3.10	1.81	1.05	0.60	8.45	4.98
14	3.82	2.25	2.67	1.59	3.10	1.81	1.05	0.60	8.45	4.98
15	3.82	2.25	2.67	1.59	3.10	1.81	1.05	0.60	8.45	4.98
16	11.45	6.76	8.00	4.78	9.31	5.44	3.16	1.81	25.34	14.94
17	11.45	6.76	8.00	4.78	9.31	5.44	3.16	1.81	25.34	14.94
18	11.45	6.76	8.00	4.78	9.31	5.44	3.16	1.81	25.34	14.94
19	1.14	0.90	0.49	0.40	1.06	0.84	0.50	0.39	2.08	1.70
20	1.14	0.90	0.49	0.40	1.06	0.84	0.50	0.39	2.08	1.70
21	1.14	0.90	0.49	0.40	1.06	0.84	0.50	0.39	2.08	1.70
22	2.28	1.81	0.98	0.80	2.13	1.68	1.00	0.78	4.16	3.39
23	2.28	1.81	0.98	0.80	2.13	1.68	1.00	0.78	4.16	3.39
24	2.28	1.81	0.98	0.80	2.13	1.68	1.00	0.78	4.16	3.39
25	6.83	5.43	2.94	2.41	6.38	5.04	3.00	2.33	12.47	10.17
26	6.83	5.43	2.94	2.41	6.38	5.04	3.00	2.33	12.47	10.17
27	6.83	5.43	2.94	2.41	6.38	5.04	3.00	2.33	12.47	10.17
28	1.14	0.90	0.49	0.40	1.06	0.84	0.50	0.39	2.08	1.70
29	1.14	0.90	0.49	0.40	1.06	0.84	0.50	0.39	2.08	1.70
30	1.14	0.90	0.49	0.40	1.06	0.84	0.50	0.39	2.08	1.70
31	2.28	1.81	0.98	0.80	2.13	1.68	1.00	0.78	4.16	3.39
32	2.28	1.81	0.98	0.80	2.13	1.68	1.00	0.78	4.16	3.39
33	2.28	1.81	0.98	0.80	2.13	1.68	1.00	0.78	4.16	3.39
34	6.83	5.43	2.94	2.41	6.38	5.04	3.00	2.33	12.47	10.17
35	6.83	5.43	2.94	2.41	6.38	5.04	3.00	2.33	12.47	10.17
36	6.83	5.43	2.94	2.41	6.38	5.04	3.00	2.33	12.47	10.17
P: Parametric, NP: Nonparametric Percentile										